Hensman Moss, Davina J., Pardinas, Antonio F., Langbehn, Douglas, Lo, Kitty, Leavitt, Blair R., Roos, Raymund, Durr, Alexandra, Mead, Simon, Holmans, Peter, Jones, Lesley, Tabrizi, Sarah J., Coleman, A., Santos, R. Dar, Decolongon, J., Sturrock, A., Bardinet, E., Ret, C. Jauff, Justo, D., Lehericy, S., Marelli, C., Nigaud, K., Valabrègue, R., van den Bogaard, S. J. A., Dumas, E. M., van der Grond, J., t'Hart, E. P., Jurgens, C., Witjes-Ane, M.-N., Arran, N., Callaghan, J., Stopford, C., Frost, C., Jones, R., Hobbs, N., Lahiri, N., Ordidge, R., Owen, G., Pepple, T., Read, J., Say, M., Wild, E., Patel, A., Fox, N. C., Gibbard, C., Malone, I., Crawford, H., Whitehead, D., Keenan, S., Cash, D. M., Berna, C., Bechtel, N., Bohlen, S., Man, A . Hoff, Kraus, P., Axelson, E., Wang, C., Acharya, T., Lee, S., Monaco, W., Campbell, C., Queller, S., Whitlock, K., Campbell, C., Campbell, M., Frajman, E., Milchman, C., O'Regan, A., Labuschagne, I., Stout, J., Landwehrmeyer, B., Craufurd, D., Scahill, R., Hicks, S., Kennard, C., Johnson, H., Tobin, A., Rosas, H. D., Reilmann, R., Borowsky, B., Pourchot, C., Andrews, S. C., Bachoud-Lévi, Anne-Catherine, Bentivoglio, Anna Rita, Biunno, Ida, Bonelli, Raphael, Burgunder, Jean-Marc, Dunnett, Stephen, Ferreira, Joaquim, Handley, Olivia, Heiberg, Arvid, Illmann, Torsten, Landwehrmeyer, G. Bernhard, Levey, Jamie, Ramos-Arroyo, Maria A., Nielsen, Jørgen, Koivisto, Susana Pro, Päivärinta, Markku, Roos, Raymund A.C., Sebastián, A. Rojo, Tabrizi, Sarah, Vandenberghe, Wim, Verellen-Dumoulin, Christine, Uhrova, Tereza, Wahlström, Jan, Zaremba, Jacek, Baake, Verena, Barth, Katrin, Garde, Monica Bascuñana, Betz, Sabrina, Bos, Reineke, Callaghan, Jenny, Come, Adrien, Guedes, Leonor Correia, Ecker, Daniel, Finisterra, Ana Maria, Fullam, Ruth, Gilling, Mette, Gustafsson, Lena, Handley, Olivia J., Hvalstedt, Carina, Held, Christine, Koppers, Kerstin, Lamanna, Claudia, Laurà, Matilde, Descals, Asunción Martínez, Martinez-Horta, Saül, Mestre, Tiago, Minster, Sara, Monza, Daniela, Mütze, Lisanne, Oehmen, Martin, Orth, Michael, Padieu, Hélène, Paterski, Laurent, Peppa, Nadia, Koivisto, Susana Pro, Di Renzo, Martina, Rialland, Amandine, Røren, Niini, ?a?inková, Pavla, Timewell, Erika, Townhill, Jenny, Cubillo, Patricia Trigo, da Silva, Wildson Vieira, van Walsem, Marleen R, Whalstedt, Carina, Witjes-Ané, Marie-Noelle, Witkowski, Grzegorz, Wright, Abigail, Zielonka, Daniel, Zielonka, Eugeniusz, Zinzi, Paola, Bonelli, Raphael M., Lilek, Sabine, Hecht, Karen, Herranhof, Brigitte, Holl, Anna, Hammer, Hans-Peter, Koppitz, Michael, Magnet, Markus, Müller, Nicole, [illegible]niela, Painold, Annamaria, Reisinger, Karin, Scheibl, Monika, Schöggl, Tamm[illegible], Ullah, Jasmin, Braunwarth, Eva-Maria, Brugger, Florian, Buratti, Lisa,

Hametner, Eva-Maria, Hepperger, Caroline, Holas, Christiane, Hotter, Anna, Hussl, Anna, Müller, Christoph, Poewe, Werner, Seppi, Klaus, Sprenger, Fabienne, Wenning, Gregor, Boogaerts, Andrea, Calmeyn, Godelinde, Delvaux, Isabelle, Liessens, Dirk, Somers, Nele, Dupuit, Michel, Minet, Cécile, van Paemel, Dominique, Ribaï, Pascale, Verellen-Dumoulin, Christine, Boogaerts, Andrea, Vandenberghe, Wim, van Reijen, Dimphna, Klempír, Jirí, Majerová, Veronika, Roth, Jan, Stárková, Irena, Hjermind, Lena E., Jacobsen, Oda, Nielsen, Jørgen E., Larsen, Ida Unmack, Vinther-Jensen, Tua, Hiivola, Heli, Hyppönen, Hannele, Martikainen, Kirsti, Tuuha, Katri, Allain, Philippe, Bonneau, Dominique, Bost, Marie, Gohier, Bénédicte, Guérid, Marie-Anne, Olivier, Audrey, Prundean, Adriana, Scherer-Gagou, Clarisse, Verny, Christophe, Babiloni, Blandine, Debruxelles, Sabrina, Duché, Charlotte, Goizet, Cyril, Jameau, Laetitia, Lafoucrière, Danielle, Spampinato, Umberto, Barthélémy, Rekha, De Bruycker, Christelle, Carette, Maryline Cabaret Anne-Sophie, Defebvre, Eric Decorte Luc, Delliaux, Marie, Delval, Arnaud, Destee, Alain, Dujardin, Kathy, Lemaire, Marie-Hélène, Manouvrier, Sylvie, Peter, Mireille, Plomhouse, Lucie, Sablonnière, Bernard, Simonin, Clémence, Thibault-Tanchou, Stéphanie, Vuillaume, Isabelle, Bellonet, Marcellin, Berrissoul, Hassan, Blin, Stéphanie, Courtin, Françoise, Duru, Cécile, Fasquel, Véronique, Godefroy, Olivier, Krystkowiak, Pierre, Mantaux, Béatrice, Roussel, Martine, Wannepain, Sandrine, Azulay, Jean-Philippe, Delfini, Marie, Eusebio, Alexandre, Fluchere, Frédérique, Mundler, Laura, Anheim, Mathieu, Julié, Celine, Boukbiza, Ouhaid Lagha, Longato, Nadine, Rudolf, Gabrielle, Tranchant, Christine, Zimmermann, Marie-Agathe, Kosinski, Christoph Michael, Milkereit, Eva, Probst, Daniela, Reetz, Kathrin, Sass, Christian, Schiefer, Johannes, Schlangen, Christiane, Werner, Cornelius J., Gelderblom, Harald, Priller, Josef, Prüß, Harald, Spruth, Eike Jakob, Ellrichmann, Gisa, Herrmann, Lennard, Hoffmann, Rainer, Kaminski, Barbara, Kotz, Peter, Prehn, Christian, Saft, Carsten, Lange, Herwig, Maiwald, Robert, Löhle, Matthias, Maass, Antonia, Schmidt, Simone, Bosredon, Cecile, Storch, Alexander, Wolz, Annett, Wolz, Martin, Capetian, Philipp, Lambeck, Johann, Zucker, Birgit, Boelmans, Kai, Ganos, Christos, Heinicke, Walburgis, Hidding, Ute, Lewerenz, Jan, Münchau, Alexander, Orth, Michael, Schmalfeld, Jenny, Stubbe, Lars, Zittel, Simone, Diercks, Gabriele, Dressler, Dirk, Gorzolla, Heike, Schrader, Christoph, Tacik, Pawel, Ribbat, Michael, Longinus, Bernhard, Bürk, Katrin, Möller, Jens Carsten, Rissling, Ida, Mühlau, Mark, Peinemann, Alexander, Städtler, Michael, Weindl, Adolf, Winkelmann, Juliane, Ziegler, Cornelia, Bechtel, Natalie, Beckmann, Heike, Bohlen, Stefan, Hölzner, Eva, Lange, Herwig, Reilmann, Ralf, Rohm, Stefanie, Rumpf, Silke, Schepers, Sigrun, Weber, Natalia, Dose, Matthias, Leythäuser, Gabriele, Marquard, Ralf, Raab, Tina, Wiedemann, Alexandra, Barth, Katrin, Buck, Andrea, Connemann, Julia, Ecker, Daniel, Geitner, Carolin, Held, Christine, Kesse, Andrea, Landwehrmeyer, Bernhard, Lang, Christina, Lewerenz, Jan, Lezius, Franziska, Nepper, Solveig, Niess, Anke, Orth, Michael, Schneider, Ariane, Schwenk, Daniela, Süßmuth, Sigurd, Trautmann, Sonja, Weydt, Patrick, Cormio, Claudia, Sciruicchio, Vittorio, Serpino, Claudia, de Tommaso, Marina, Capellari, Sabina, Cortelli, Pietro, Galassi, Roberto, Rizzo, Giovanni, Poda, Roberto, Scaglione, Cesa, Bertini, Elisabetta, Ghelli, Elena, Ginestroni, Andrea, Massaro, Francesca, Mechi, Claudia, Paganini, Marco, Piacentini, Silvia, Pradella, Silvia, Romoli, Anna Maria, Sorbi, Sandro, Abbruzzese, Giovanni, di Poggio, Monica Bandettini, Ferrandes, Giovanna, Mandich, Paola, Marchese, Roberta, Albanese, Alberto, Di Bella, Daniela, Castaldo, Anna, Di Donato,

Stefano, Gellera, Cinzia, Genitrini, Silvia, Mariotti, Caterina, Monza, Daniela, Nanetti, Lorenzo, Paridi, Dominga, Soliveri, Paola, Tomasello, Chiara, De Michele, Giuseppe, Di Maio, Luigi, Massarelli, Marco, Peluso, Silvio, Roca, Alessandro, Russo, Cinzia Valeria, Salvatore, Elena, Sorrentino, Pierpaolo, Amico, Enrico, Favellato, Mariagrazia, Griguoli, Annamaria, Mazzante, Irene, Petrollini, Martina, Squitieri, Ferdinando, D'Alessio, Barbara, Esposito, Chiara, Bentivoglio, Rita, Frontali, Marina, Guidubaldi, Arianna, Ialongo, Tamara, Jacopini, Gioia, Piano, Carla, Romano, Silvia, Soleti, Francesco, Spadaro, Maria, Zinzi, Paola, van Hout, Monique S.E., Verhoeven, Marloes E., van Vugt, Jeroen P.P., de Weert, A. Marit, Bolwijn, J.J.W., Dekker, M., Kremer, B., Leenders, K.L., van Oostrom, J.C.H., van den Bogaard, Simon J.A., Bos, Reineke, Dumas, Eve M., 't Hart, Ellen P., Roos, Raymund A.C., Kremer, Berry, Verstappen, C.C.P., Aaserud, Olaf, C, Jan Frich, Heiberg, Arvid, van Walsem, Marleen R, Wehus, Ragnhild, Bjørgo, Kathrine, Fannemel, Madeleine, Gørvell, Per F., Lorentzen, Eirin, Koivisto, Susana Pro, Retterstøl, Lars, Stokke, Bodil, Bjørnevoll, Inga, Sando, Sigrid Botne, Dziadkiewicz, Artur, Nowak, Malgorzata, Robowski, Piotr, Sitek, Emilia, Slawek, Jaroslaw, Soltan, Witold, Szinwelski, Michal, Blaszcyk, Magdalena, Boczarska-Jedynak, Magdalena, Ciach-Wysocka, Ewelina, Gorzkowska, Agnieszka, Jasinska-Myga, Barbara, Klodowska-Duda, Gabriela, Opala, Gregorz, Stompel, Daniel, Banaszkiewicz, Krzysztof, Bocwinska, Dorota, Bojakowska-Jaremek, Kamila, Dec, Malgorzata, Krawczyk, Malgorzata, Rudzinska, Monika, Szczygiel, Elzbieta, Szczudlik, Andrzej, Wasielewska, Anna, Wójcik, Magdalena, Bryl, Anna, Ciesielska, Anna, Klimberg, Aneta, Marcinkowski, Jerzy, Samara, Husam, Sempolowicz, Justyna, Zielonka, Daniel, Gogol, Anna, Janik, Piotr, Kwiecinski, Hubert, Jamrozik, Zygmunt, Antczak, Jakub, Jachinska, Katarzyna, Krysa, Wioletta, Rakowicz, Maryla, Richter, Przemyslaw, Rola, Rafal, Ryglewicz, Danuta, Sienkiewicz-Jarosz, Halina, Stepniak, Iwona, Sulek, Anna, Witkowski, Grzegorz, Zaremba, Jacek, Zdzienicka, Elzbieta, Zieora-Jakutowicz, Karolina, Ferreira, Joaquim J., Coelho, Miguel, Guedes, Leonor Correia, Mendes, Tiago, Mestre, Tiago, Valadas, Anabela, Andrade, Carlos, Gago, Miguel, Garrett, Carolina, Guerra, Maria Rosália, Herrera, Carmen Durán, Garcia, Patrocinio Moreno, Barbera, Miquel Aguilar, Guia, Dolors Badenes, Hernanz, Laura Casas, Catena, Judit López, Ferrer, Pilar Quiléz, Sebastián, Ana Rojo, Carruesco, Gemma Tome, Bas, Jordi, Busquets, Núria, Calopa, Matilde, Robert, Misericordia Floriach, Viladrich, Celia Mareca, Idiago, Jesús Miguel Ruiz, Riballo, Antonio Villa, Cubo, Esther, Polo, Cecilia Gil, Mariscal, Natividad, Rivadeneyra, Perez Jessica, Barrero, Francisco, Morales, Blas, Fenollar, María, García, Rocío García-Ramos, Ortega, Paloma, Villanueva, Clara, Alegre, Javier, Bascuñana, Mónica, Caldentey, Juan Garcia, Ventura, Marta Fatás, Ribas, Guillermo García, de Yébenes, Justo García, Moreno, José Luis López-Sendón, Cubillo, Patricia Trigo, Alegre, Javier, Frech, Fernando Alonso, de Yébenes, Justo García, Ruíz, Pedro J García, Martínez-Descals, Asunción, Guerrero, Rosa, Artiga, María José Saiz, Sánchez, Vicenta, Perea, María Fuensanta Noguera, Fortuna, Lorenza, Manzanares, Salvadora, Reinante, Gema, Torres, María Martirio Antequera, Moreau, Laura Vivancos, González González, Sonia, Guisasola, Luis Menéndez, Salvador, Carlos, Martín, Esther Suaréz San, Ramirez, Inés Legarda, Gorospe, Aranzazú, Lopera, Mónica Rodriguez, Arques, Penelope Navas, Rodríguez, María José Torres, Pastor, Barbara Vives, Gaston, Itziar, Martinez-Jaurrieta, Maria Dolores, Ramos-Arroyo, Maria A., Moreno, Jose Manuel Garcia, Lucena, Carolina Mendez, Damas, Fatima, Cortegana, Hermoso Eva Pacheco, Peña, José Chacón, Redondo, Luis, Carrillo,

Fátima, Teresa Cáceres, María, Mir, Pablo, Suarez, María José Lama, Vargas-González, Laura, Bosca, Maria E., Brugada, Francisco Castera, Burguera, Juan Andres, Campos, Anabel, Vilaplana, Garcia Carmen Peiró, Berglund, Peter, Constantinescu, Radu, Fredlund, Gunnel, Høsterey-Ugander, Ulrika, Linnsand, Petra, Neleborn-Lingefjärd, Liselotte, Wahlström, Jan, Wentzel, Magnus, Loutfi, Ghada, Olofsson, Carina, Stattin, Eva-Lena, Westman, Laila, Wikström, Birgitta, Burgunder, Jean-Marc, Stebler, Yanik, Kaelin, Alain, Romero, Irene, Schüpbach, Michael, Weber Zaugg, Sabine, Hauer, Maria, Gonzenbach, Roman, Jung, Hans H., Mihaylova, Violeta, Petersen, Jens, Jack, Roisin, Matheson, Kirsty, Miedzybrodzka, Zosia, Rae, Daniela, Simpson, Sheila A, Summers, Fiona, Ure, Alexandra, Vaughan, Vivien, Akhtar, Shahbana, Crooks, Jenny, Curtis, Adrienne, de Souza, Jenny, Piedad, John, Rickards, Hugh, Wright, Jan, Coulthard, Elizabeth, Gethin, Louise, Hayward, Beverley, Sieradzan, Kasia, Wright, Abigail, Armstrong, Matthew, Barker, Roger A., O'Keefe, Deidre, Di Pietro, Anna, Fisher, Kate, Goodman, Anna, Hill, Susan, Kershaw, Ann, Mason, Sarah, Paterson, Nicole, Raymond, Lucy, Swain, Rachel, Guzman, Natalie Valle, Busse, Monica, Butcher, Cynthia, Clenaghan, Catherine, Fullam, Ruth, Hunt, Sarah, Jones, Lesley, Jones, Una, Khalil, Hanan, Owen, Michael, Price, Kathleen, Rosser, Anne, Edwards, Maureen, Ho, Carrie, Hughes, Teresa, McGill, Marie, Pearson, Pauline, Porteous, Mary, Smith, Paul, Brockie, Peter, Foster, Jillian, Johns, Nicola, McKenzie, Sue, Rothery, Jean, Thomas, Gareth, Yates, Shona, Burrows, Liz, Chu, Carol, Fletcher, Amy, Gallantrae, Deena, Hamer, Stephanie, Harding, Alison, Klöppel, Stefan, Kraus, Alison, Laver, Fiona, Lewis, Monica, Longthorpe, Mandy, Markova, Ivana, Raman, Ashok, Robertson, Nicola, Silva, Mark, Thomson, Aileen, Wild, Sue, Yardumian, Pam, Chu, Carol, Evans, Carole, Gallentrae, Deena, Hamer, Stephanie, Kraus, Alison, Raman, Ashok, Chu, Carol, Hamer, Stephanie, Hobson, Emma, Jamieson, Stuart, Raman, Ashok, Musgrave, Hannah, Rowett, Liz, Toscano, Jean, Bourne, Colin, Clapton, Jackie, Clayton, Carole, Dipple, Heather, Freire-Patino, Dawn, Grant, Janet, Gross, Diana, Hallam, Caroline, Middleton, Julia, Murch, Ann, Thompson, Catherine, Alusi, Sundus, Davies, Rhys, Foy, Kevin, Gerrans, Emily, Pate, Louise, Andrews, Thomasin, Dougherty, Andrew, Golding, Charlotte, Kavalier, Fred, Laing, Hana, Lashwood, Alison, Robertson, Dene, Ruddy, Deborah, Santhouse, Alastair, Whaite, Anna, Andrews, Thomasin, Bruno, Stefania, Doherty, Karen, Golding, Charlotte, Haider, Salman, Hensman, Davina, Lahiri, Nayana, Lewis, Monica, Novak, Marianne, Patel, Aakta, Rosser, Elisabeth, Tabrizi, Sarah, Taylor, Rachel, Warner, Thomas, Wild, Edward, Arran, Natalie, Bek, Judith, Craufurd, David, Fullam, Ruth, Hare, Marianne, Howard, Liz, Huson, Susan, Johnson, Liz, Jones, Mary, Murphy, Helen, Oughton, Emma, Partington-Jones, Lucy, Rogers, Dawn, Sollom, Andrea, Snowden, Julie, Stopford, Cheryl, Thompson, Jennifer, Trender-Gerhard, Iris, Verstraelen, Nichola, Westmoreland, Leann, Armstrong, Richard, Dixon, Kathryn, Nemeth, Andrea H., Siuda, Gill, Valentine, Ruth, Harrison, David, Hughes, Max, Parkinson, Andrew, Soltysiak, Beverley, Bandmann, Oliver, Bradbury, Alyson, Gill, Paul, Fairtlough, Helen, Fillingham, Kay, Foustanos, Isabella, Kazoka, Mbombe, O'Donovan, Kirsty, Peppa, Nadia, Taylor, Cat, Tidswell, Katherine, Quarrell, Oliver, Burgunder, Jean-Marc, Lau, Puay Ngoh, Pica, Emmanul and Tan, Louis 2017. Identification of genetic variants associated with Huntington's disease progression: a genome-wide association study. Lancet Neurology 16 (9) , pp. 701-711. 10.1016/S1474-4422(17)30161-8 file

Identification of genetic variants associated with Huntington's disease progression: a genome-wide association study

Davina J Hensman Moss*[1], MBBS, Antonio F. Pardiñas*[2], PhD, Prof Douglas Langbehn[3], PhD, Kitty Lo[4], PhD, Prof Blair R. Leavitt[5], MD,CM, Prof Raymund Roos[6], MD, Prof Alexandra Durr[7], MD, Prof Simon Mead[8], PhD, the REGISTRY investigators and the TRACK-HD investigators, Prof Peter Holmans[2], PhD, Prof Lesley Jones[§2], PhD, Prof Sarah J Tabrizi[§1], PhD.

* These authors contributed equally to this work

§ These authors contributed equally to this work

1) UCL Huntington's Disease Centre, UCL Institute of Neurology, Dept. of Neurodegenerative Disease, London, UK
2) MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, UK
3) University of Iowa Carver College of Medicine, Dept. of Psychiatry and Biostatistics, Iowa, USA
4) UCL Genetics Institute, Div. of Biosciences, London, UK
5) Centre for Molecular Medicine and Therapeutics, Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada
6) Department of Neurology, Leiden University Medical Centre, Leiden, Netherlands
7) ICM and APHP Department of Genetics, Inserm U 1127, CNRS UMR 7225, Sorbonne Universités, UPMC Univ Paris 06 UMR S 1127, Pitié- Salpêtrière University Hospital, Paris, France
8) MRC Prion Unit, UCL Institute of Neurology, London, UK

Corresponding authors:

Sarah J Tabrizi at s.tabrizi@ucl.ac.uk

Lesley Jones at JonesL1@cardiff.ac.uk

ABSTRACT

Background Huntington's disease (HD) is a fatal inherited neurodegenerative disease, caused by a CAG repeat expansion in *HTT*. Age at onset (AAO) has been used as a quantitative phenotype in genetic analysis looking for HD modifiers, but is hard to define and not always available. Therefore here we aimed to generate a novel measure of disease progression, and identify genetic markers associated with this progression measure.

Methods We generated a progression score based on principal component analysis of prospectively acquired longitudinal changes in motor, behavioural, cognitive and imaging measures in the TRACK-HD cohort of HD gene mutation carriers (data collected 2008 – 2011). We generated a parallel progression score using 1773 previously genotyped subjects from the REGISTRY study of HD mutation carriers (data collected 2003 – 2013). 216 subjects from TRACK-HD were genotyped. Association analyses was performed using GCTA, gene-wide analysis using MAGMA and meta-analysis using METAL.

Findings Longitudinal motor, cognitive and imaging scores were correlated with each other in TRACK-HD subjects, justifying a single, cross-domain measure as a unified progression measure in both studies. The TRACK-HD and REGISTRY progression measures were correlated with each other (r=0·674), and with AAO (r=0·315, r=0.234 respectively). A meta-analysis of progression in TRACK-HD and REGISTRY gave a genome-wide significant signal ($p=1.12x10^{-10}$) on chromosome 5 spanning 3 genes, *MSH3, DHFR* and *MTRNR2L2*. The lead SNP in TRACK-HD (rs557874766) is genome-wide significant in the meta-analysis ($p=1.58x10^{-8}$), and encodes an amino acid change (Pro67Ala) in MSH3. In TRACK-HD, each copy of the minor allele at this SNP is associated with a 0.4 (95% CI=0.16,0.66) units per year reduction in the rate of change of the Unified Huntington's Disease Rating Scale (UHDRS) Total Motor Score, and 0.12 (95% CI=0.06,0.18) units per year in

the rate of change of UHDRS Total Functional Capacity. The associations remained significant after adjusting for AAO.

Interpretation The multi-domain progression measure in TRACK-HD is associated with a functional variant that is genome-wide significant in a meta-analysis. The strong association in only 216 subjects implies that the progression measure is a sensitive reflection of disease burden, that the effect size at this locus is large, or both. As knock out of Msh3 reduces somatic expansion in HD mouse models, this highlights somatic expansion as a potential pathogenic modulator, informing therapeutic development in this untreatable disease.

Funding sources The European Commission FP7 NeurOmics project; CHDI Foundation; the Medical Research Council UK, the Brain Research Trust, the Guarantors of Brain.

Research in context

Evidence before this study

Huntington's disease (HD) is universally caused by a tract of 36 or more CAG in exon 1 of *HTT*. Genetic modifiers of age at motor onset have recently been identified in HD that highlight pathways, which if modulated in people, might delay disease onset. Onset of disease is preceded by a long prodromal phase accompanied by substantial brain cell death and age at motor onset is difficult to assess accurately and is not available in disease free at risk subjects. We searched all of PubMed up to Oct 31st 2016 for articles published in English containing "Huntington* disease" AND "genetic modifier" AND "onset" which identified 13 studies, then "Huntington* disease" AND "genetic modifier" AND "progression" which identified one review article. Amongst the 13 studies of genetic modification of HD onset most were small candidate gene studies; these were superseded by the one large genome wide genetic modifiers of HD study which identified three genome-wide significant loci, and implicated DNA handling in HD disease modification

Added value of this study

We examined the prospective data from TRACK-HD and developed a measure of disease progression that reflected correlated progression in the brain imaging, motor and cognitive symptom domains: there is substantial correlation among these variables. We used the disease progression measure as a quantitative variable in a genome-wide association study and in only 216 people from TRACK-HD detected a locus on chromosome 5 containing three significant genes, *MTRNR2L2*, *MSH3* and *DHFR*. The index variant encodes an amino acid change in MSH3. We replicated this finding by generating a parallel progression measure in the less intensively phenotyped REGISTRY study and detected a similar signal on chromosome 5, likely attributable to the same variants. A meta-analysis of the two studies strengthened the associations. There was some correlation between the progression measures and AAO of disease but this was not responsible for the association with disease progression. We also detected a signal on chromosome 15 in the REGISTRY study at the same locus as that previously associated with AAO.

Implications of all the available evidence

The progression measures used in this study can be generated in asymptomatic and symptomatic subjects using a subset of the clinically relevant parameters gathered in TRACK-HD. We use these measures to identify genetic modifiers of disease progression in HD. We saw a signal in only 216 subjects, which replicates in a larger sample, becoming genome-wide significant, thus reducing the chance of it being a false positive. This argues for the power of better phenotypic measures in genetic studies and implies that this locus has a large effect size on disease progression. The index associated genetic variant in TRACK-HD encodes a Pro67Ala change in MSH3, which implicates *MSH3* as the associated gene on chromosome 5. Notably, altering levels of Msh3 in HD mice reduces somatic instability and crossing *Msh3* null mice with HD mouse models prevents somatic instability of the *HTT* CAG repeat and reduces pathological phenotypes. Polymorphism in *MSH3* has been linked to somatic instability in myotonic dystrophy type 1 patients. MSH3 is a non-essential

neuronally expressed member of the DNA mismatch repair pathway and these data reinforce its candidacy as a therapeutic target in HD and potentially in other neurodegenerative expanded repeat disorders.

INTRODUCTION

Huntington's disease (HD) is a autosomal dominant fatal neurodegenerative condition caused by a CAG repeat expansion in *HTT* (1). It is a movement, cognitive and psychiatric disorder, but symptoms, age of disease onset (AAO) and disease progression vary (2). AAO (1, 3) reflects the trajectory of disease pathology up to the point of motor onset. However, the transition from premanifest to manifest HD is gradual (4, 5), making clinical definition challenging, furthermore psychiatric and cognitive changes may not be concurrent with motor onset (6). Despite this imprecision in defining onset, the inverse correlation of *HTT* CAG repeat length and age at motor onset accounts for 50-70% of the observed variance in onset (7). Part of the remaining difference in onset age was recently shown to be genetically encoded, identifying genes of the DNA damage response as likely to modify onset of HD (8).

The need for clinical trials close to disease onset has motivated a raft of observational studies (5, 9, 10). This provides the opportunity to investigate the relationship between onset and progression, whether they are influenced by the same biology, and permits the study of subjects before clinical onset.

TRACK-HD represents the most deeply phenotyped cohort of premanifest and symptomatic disease with annual visits involving clinical, cognitive and motor testing alongside detailed brain imaging (5, 6). We used TRACK-HD (5, 6) data to generate a novel unified Huntington's disease progression measure for use in a genetic association analysis. We developed a similar measure in subjects from the REGISTRY study to replicate our findings (9).

MATERIALS AND METHODS

Study design and participants

All experiments were performed in accordance with the Declaration of Helsinki and approved by the University College London (UCL)/UCL Hospitals Joint Research Ethics Committee; ethical approval for the REGISTRY analysis is outlined in (8). Peripheral blood samples were donated by genetically-confirmed HD gene carriers, and all subjects provided informed written consent.

TRACK-HD was a prospective observational biomarker study collecting deep phenotypic data including imaging, quantitative motor and cognitive assessments on adult subjects with early HD, premanifest HD gene carriers and controls (5, 6). It provides annually collected high quality longitudinal prospective multivariate data over three years (2008-2011) with 243 subjects at baseline (6) (**Figure 1**). Demographic details of these individuals are shown in **Supplementary Information**.

REGISTRY(9) was a multisite prospective observational study which collected phenotypic data between 2003 – 2013 on over 13,000 subjects, mostly manifest HD gene carriers. The aim is for annual assessments +/- 3 months, though this is variable. The core data include: age, CAG repeat length, UHDRS Total Motor Score (TMS) and Total Functional Capacity (TFC); some patients have further assessments such as a cognitive battery (9). 1835 adult subjects from REGISTRY were included in this study on the basis of available genotype data (8). We obtained: TMS, symbol digit modality (SDMT), verbal fluency, Stroop colour reading, word reading and interference measures, functional assessment score, and TFC.

Procedures

For both studies, atypical severity scores were derived with a combination of principal component analysis (PCA) and regression of the predictable effects of the primary gene *HTT* CAG repeat length.

Details differed however, due to differences in nature of the two data sets. In TRACK-HD, 24 variables were used to stratify the cohort in terms of disease progression (**Supplementary Information)**. They were divided *a priori* into 3 broad domains: (1) brain volume measures, (2) cognitive variables, and (3) quantitative-motor variables. For each variable the input for analysis was the subject's random longitudinal slope from a mixed effects regression model with correlated random intercepts and slopes for each subject. This model regressed the observed values on clinical probability of onset statistic (CPO) derived from CAG repeat length and age, and its interaction with follow-up length. The subjects' random slope estimates thus provided a measure of atypical longitudinal change not predicted by age and CAG length. Principal Component Analyses (PCA) of the random slopes was then used to study the dimensionality of these age and CAG-length corrected longitudinal changes. Further methodological detail, including control for potential demographic confounders, is given in Supplementary Methods and a flow chart is given in **Figure 1**.

For REGISTRY, in contrast to TRACK-HD, follow-up length and frequency was variable and missing data were substantial, making longitudinal progression analysis problematic. We therefore examined cross-sectional status at last visit, using a single unified motor-cognitive dimension of severity. We performed multiple imputation to fill in missing data, derived PCA severity scores and regressed off the predictive effect of age, CAG length, and gender on the PCA severity scores derived from this data to obtain the measure of atypical severity at the last visit. This gives a single point "severity" score based on how advanced a subject is compared with expectations based on their CAG repeat and age. 1773 subjects had adequate phenotypic data to score; further detail is given in Supplementary Methods and a flow chart is given in **Figure 1.**

Statistical and genetic analysis

Data analyses were performed using SAS/STAT 14·0 and 14·1 primarily via the MIXED, FACTOR and GML procedures (11). We occasionally used a log or inverse transform of a measure, with the

goal of better approximate normality of the distribution and the avoidance of inappropriate influence of extreme scores.

218 TRACK-HD study participants with complete serial phenotype data were genotyped on Illumina Omni2.5v1·1 arrays, and quality control performed as described in Supplementary Methods. Imputation was carried out using the 1000 Genomes phase 3 data as a reference (Supplementary Methods). This yielded 9·65 million biallelic markers of 216 individuals. Genotypes for the REGISTRY subjects were obtained from the GeM-HD Consortium (8), where details of their genotyping, quality control, curation and imputation are provided.

Association analyses were performed with the mixed linear model (MLM) functions included in GCTA v1·26(12). Conditional analyses were carried out using the COJO procedure included in GCTA. Because of the relatively small sample sizes, analyses were restricted to SNPs with minor allele frequency >1%. A meta-analysis of the TRACK-HD and REGISTRY association results was performed using METAL(13). To test whether the association signals in TRACK-HD and REGISTRY could have arisen from the same causal SNPs, and whether these also influenced expression co-localisation analysis was carried out using GWAS-pw v0·21 (14). Gene-wide p-values were calculated using MAGMA v1·05, a powerful alternative to SNP-based analyses which aggregates the association signal inside genes while taking linkage disequilibrium (LD) between SNPs into account (15), using a window of 35kb upstream and 10kb downstream of genes (16). Such an analysis can increase power over single-SNP analysis when there are multiple causal SNPs in a gene, or when the causal SNP is not typed and its signal is partially captured by multiple typed SNPs in LD with it. To maximise comparability with the GeM GWAS, our primary pathway analyses used Setscreen (17), which sums the log p-values of all SNPs in a pathway, also correcting for LD between SNPs.

All of the methods and analyses mentioned in this section are described in more detail in Supplementary Information.

RESULTS

We performed individual PCA of each domain and found that first PC scores were highly correlated between the domains ($P < 0{\cdot}0001$ in all cases, **Supplementary Information**.) No phenotypic subtypes of symptom clusters in motor, cognitive or imaging domains were observed; rather, longitudinal change in TRACK-HD not predictable by CAG-age was distributed on a correlated continuum (**Figure 2**). We therefore repeated PCA of the measures combined across all domains. The first PC of this combined analysis accounted for 23.4% of the joint variance, and was at least moderately correlated ($r>0{\cdot}4$) with most of the variables that contributed heavily to each domain-specific first PC (**Supplementary Tables 3 and 4**). The first psychiatric PC has notably lower correlation with motor and cognitive domains and CPO variables, so was excluded from our progression measures.

The cross-domain first principal component was used as a unified Huntington's disease progression measure in the TRACK-HD cohort (**Figure 1 and 2B**). To confirm that our progression measure correlated with commonly recognised measures of Huntington's disease severity not included in the progression analysis, we examined the residual change relationships between the progression score and UHDRS TMS change and TFC change after controlling for the CPO. We found a correlation of $r=0{\cdot}448$ ($p<0{\cdot}0001$) for the residual motor slope and $r=-0{\cdot}421$ ($p<0{\cdot}0001$) for the residual TFC slope. One unit increase in unified Huntington's disease progression measure corresponded to an increase of 0·71 (95% CI=0.34,1.08) units per year in the rate of change of TMS, and an increase of approximately 0·2 (95% CI=0.12,0.30) units per year in the rate of change of TFC. The 15 fastest progressing subjects in TRACK-HD showed a mean annual rate of decline in the UHDRS TMS of 2·52 more points per year than would be expected (Standard deviation =2.47, Standard Error of Mean =0.64); the 15 slowest progressing subjects had an annual TMS decline of 0·45 points less per

year than predicted by age and CAG length (Standard deviation =1.85, Standard Error of the Mean =0.48).

Huntington's disease subjects in the early stages of the disease were significantly faster progressors on the unified HD progression measure than those still in the premanifest phase ($p < 0{\cdot}0001$). Amongst the 96 subjects who had experienced onset, the rater AAO showed the expected relation with predicted AAO based on CAG length (**Supplementary Information**), and earlier than predicted AAO was correlated with faster progression on our unified HD progression measure ($r=0{\cdot}315$; $p = 0{\cdot}002$).

The unified HD progression measure developed in TRACK-HD could not be transferred directly to REGISTRY subjects with more limited data. Individual clinical measures in REGISTRY showed correlations across the motor, cognitive, and functional domains, consistent with our finding in TRACK-HD (**Supplementary Information**). PC1 accounted for 75·6% of the variance in severity; no other principal components explained any substantial amount of the common variance within the measures used (**Supplementary Information**). Therefore this first principal component was chosen as a measure of severity in the REGISTRY cohort (**Figure 2C**). Higher values of this measure mean greater severity than expected at a given time: we infer that this is the result of faster progression (**Figure 2A**) and we used this as the unified Registry progression measure. The unified REGISTRY progression measure and earlier than predicted AAO were modestly, but significantly, correlated ($r = 0{\cdot}2338$; $p<0{\cdot}0001$) (**Supplementary Information**). Atypically rapidly or slowly progressing subjects tend to become more atypical over time: correlation between time since disease onset and REGISTRY progression ($-0{\cdot}3074$; $p<0{\cdot}0001$) is greater than that between AAO and REGISTRY progression.

In TRACK-HD, the last-visit severity scores had a correlation of 0·674 with the previously calculated longitudinal unified progression measure, indicating that our progression measures for TRACK-HD and REGISTRY reflected strongly, although not perfectly, related elements of clinical

phenotype. Further support for this conclusion was given by the correlation of 0·631 between the TRACK-HD and REGISTRY progression measures in the 14 subjects present in both studies.

We then performed a genome-wide association analysis using the unified TRACK-HD progression measure as a quantitative trait, which yielded a significantly associated locus on chromosome 5 spanning *DHFR*, *MSH3* and *MTRNR2L2*. The index SNP rs557874766 is a coding missense variant in *MSH3* ($p = 5{\cdot}8 \times 10^{-8}$; G=0·2179/1091 (1000 Genomes); **Figure 3A and D and Supplementary Information**). Analyses conditioning on this SNP failed to show evidence for a second independent signal in this region in TRACK-HD **(Supplementary Information).** The genes in this locus were the only ones to reach genome-wide genic significance ((15, 18) (*MTRNR2L2* $p = 2{\cdot}15 \times 10^{-9}$; *MSH3* $p = 2{\cdot}94 \times 10^{-8}$; *DHFR* $p = 8{\cdot}37 \times 10^{-7}$, **http://hdresearch.ucl.ac.uk/data-resources/**).

Performing a genome-wide association analysis in REGISTRY using the unified progression measure replicated the signal identified in TRACK-HD (lead SNP rs420522, $p = 1{\cdot}39 \times 10^{-5}$) on a narrower locus (chr5:79902336-79950781), but still tagging the same three genes (**Figure 3B and D**). No genes reach genome-wide significance, though there is evidence of association (**http://hdresearch.ucl.ac.uk/data-resources/)** at *DHFR* ($p = 8{\cdot}45 \times 10^{-4}$), *MSH3* ($p = 9{\cdot}36 \times 10^{-4}$), and *MTRNR2L2* ($p = 1{\cdot}20 \times 10^{-3}$).

The meta-analysis of TRACK-HD and REGISTRY strengthened the signal of both individual SNPs in this region, encompassing the first three exons of *MSH3* along with *DHFR* and *MTRNR2L2* (**Figure 4C and D, Supplementary Information**), and also genic associations over *MSH3*, *DHFR*, and *MTRNR2L2* (**http://hdresearch.ucl.ac.uk/data-resources/**). The most significant SNP in the meta-analysis is rs1232027, which is genome-wide significant ($p = 1.12 \times 10^{-10}$), with the p-value of rs557874766 being 1.58×10^{-8}. No other regions attained genome-wide significance **(http://hdresearch.ucl.ac.uk/data-resources/).** Rs557874766 is nominally significant in REGISTRY (p=0.010), with a direction of effect consistent with that in TRACK-HD. Analyses conditional on rs1232027 largely remove the association in this region **(Supplementary**

Information), suggesting that there is only one signal. Conditioning on rs557874766 has a similar effect **(Supplementary Information)**, so this SNP remains a plausible causal variant.

As suggested by the meta-analysis, co-localisation analyses between TRACK-HD and REGISTRY showed this locus was likely influenced by the same SNPs in both studies (posterior probability 74.33%), although conditioning REGISTRY on rs55787466 did not remove the association signal entirely **(Supplementary Information)**. Co-localisation analyses with the GTeX expression data (19) showed strong evidence (posterior probability 96-99%) that SNPs influencing progression in TRACK-HD were also eQTLs for DHFR in brain and peripheral tissues (**Supplementary Information**). Conversely, there was strong evidence (posterior probability=97·8%) that progression SNPs in REGISTRY were eQTLs for MSH3 in blood and fibroblasts (**Supplementary Information**). Despite the lack of co-localisation between the TRACK GWAS and MSH3 expression signal, several of the most significant GWAS SNPs were associated with decreased MSH3 expression and slower progression (**Supplementary Information**). Thus, the signal on chromosome 5 could be due to the coding change in *MSH3*, or to expression changes in MSH3, DHFR or both, and both effects may operate in disease.

The second most significant association region in REGISTRY (**Supplementary Information**) tags a locus on chromosome 15 which has been previously associated to HD AAO (8). Five genes were highlighted, two of which reached genome-wide genic significance (*MTMR10* $p=2{\cdot}51x10^{-7}$; *FAN1* $p=2{\cdot}35x10^{-6}$, **http://hdresearch.ucl.ac.uk/data-resources/**). Notably, *MLH1* on chr3 contains SNPs approaching genome-wide significance ($p = 2.2 \times 10^{-7}$) in GeM-HD (8), and also shows association in the REGISTRY progression gene-wide analysis ($p = 3{\cdot}97x10^{-4}$).

As noted earlier, both progression measures are correlated with AAO. Thus, to test whether there is an association with progression independent of AAO, we repeated the REGISTRY progression GWAS conditioning for the AAO measure previously associated with this locus in GeM in the individuals (N=1,314) for whom we had measures of both progression and AAO. Both *MTMR10*

($p=1{\cdot}33x10^{-5}$) and *FAN1* ($p=1{\cdot}68x10^{-4}$) remained significant (**http://hdresearch.ucl.ac.uk/data-resources/**). Furthermore, the most significant SNP (rs10611148, $p=2{\cdot}84x10^{-7}$) was still significant after conditioning on AAO ($p=2{\cdot}40x10^{-5}$). Notably, the genic associations at the *MSH3* locus in the TRACK-HD sample also remain significant after correcting for AAO (**http://hdresearch.ucl.ac.uk/data-resources/),** as does the association with rs557874766 ($p=6{\cdot}30x10^{-6}$). A similar pattern is observed at the *MSH3* locus in the meta-analysis. Thus, the associations reported here are mainly due to disease progression, rather than AAO.

Gene set analysis of the 14 pathways highlighted by the GeM-HD paper (8) show that the four most significant pathways in the TRACK-HD progression GWAS are related to mismatch repair, and all show significant enrichment of signal in REGISTRY (**Table 1**). This enrichment is strengthened in the meta-analysis (**Table 1**). Notably, the top two pathways in TRACK-HD are also significant in the MAGMA competitive gene-set analysis (GO:32300 p=0·010, KEGG:3430 p=0·00697). *MSH3* ($2.94x10^{-8}$) and *POLD2* ($7{\cdot}21x10^{-4}$) show association in TRACK, with *MSH3* ($9{\cdot}52x10^{-4}$) and *MLH1* ($3{\cdot}97x10^{-4}$) showing association in REGISTRY (**Supplementary Information).** These findings are supported by analysis of DNA damage response pathways derived from Pearl *et al*. (20) (**Figure 4A, Supplementary Information)** where two mismatch repair pathways are significantly associated with the unified TRACK-HD progression measure after correction for multiple testing of pathways. Again, the meta-analysis strengthens the enrichment **(Figure 4B, Supplementary Information**). Genes from the two significant pathways in TRACK-HD are shown in the **Supplementary Information**, with the significant genes being very similar to those from the GeM pathways (**Supplementary Information).** A complete list of genes in the Pearl et al. (20) pathways is given in **http://hdresearch.ucl.ac.uk/data-resources/.**

DISCUSSION

The evidence from our study suggests that *MSH3* is likely to be a modifier of disease progression in Huntington's disease. We undertook an unbiased genetic screen using a novel disease progression measure in the TRACK-HD study, and identified a significant locus on chromosome 5, which encompasses three genes: *MTRNR2L2, MSH3* and *DHFR*. This locus replicated in an independent group of subjects from the European HD REGISTRY study using a parallel disease progression measure, and was genome-wide significant in a meta-analysis of the two studies. The lead SNP in TRACK-HD, rs557874766, is a coding variant in MSH3; it is classed of moderate impact, making it genome-wide significant given its annotation (21). This SNP becomes clearly genome-wide significant at the more widely used threshold of $p=5x10^{-8}$ in a meta-analysis of TRACK-HD and REGISTRY. Furthermore, eQTL analyses show association of lower MSH3 expression with slower disease progression.

Genetic modifiers of disease in people highlight pathways for therapeutic development; any pathway containing genetic variation that ameliorates or exacerbates disease forms a pre-validated relevant target. However, while the classical case-control design in complex disease has yielded multiple genetic associations highlighting relevant biology for novel treatment design (22), studies of potential genetic modifiers in genetically simple Mendelian diseases have been difficult to conduct. The diseases are rare and show gene and locus heterogeneity, thus finding genuine modifying associations in such a noisy background is inherently difficult. However, variants that modify disease in the context of a Mendelian causative gene may not be under negative selection pressure in the general population. Recent successful identifications of modifiers have been made in specific genetic subtypes of disease (23) or in relatively large samples with consistent clinical data (8, 24).

One way to increase the power of genetic studies is to obtain a more accurate measure of phenotype. Prospective multivariate longitudinal measures such as those collected in TRACK-HD are ideal (25). Our analysis of Huntington's disease progression showed that motor, cognitive and brain imaging variables typically progress in parallel and that patterns of loss are not sufficiently distinct to be

considered sub-phenotypes for genetic analysis. As psychiatric symptoms showed a different trajectory, we developed a single progression measure excluding the psychiatric data (**Figure 2A and B**). AAO was correlated with the unified progression measure but did not explain the genetic associations observed with progression. Thus, progression seems to be measuring a different aspect of disease to AAO, or a similar aspect of disease, but with greater precision. The data available in REGISTRY are less comprehensive; therefore we used a different approach by comparing cross-sectional severity at the most recent visit with that expected based on age and CAG. The unified progression measures in TRACK-HD and REGISTRY are correlated and again, the genetic associations in REGISTRY are not completely driven by AAO, demonstrating the utility of retrospective composite progression scores in genetic analysis. Prognostic indices for motor onset have been developed (26), and the development of progression scores for prospective use, for example to empower drug trials by stratifying patients by predicted rate of progression warrants further attention.

However, our study has a number of limitations. TRACK-HD has the same standardised detailed phenotypic information on nearly all participants, but in only 243 HD gene mutation carrying subjects. The REGISTRY study is much larger but the phenotypic data are less complete (**Supplementary Information**), often not collected at regular intervals and not on everyone in the study, and in multiple centres which will inevitably lead to intrinsic variation. Nevertheless, the progression measures show the expected relationship with change in TMS and TFC in both TRACK-HD and REGISTRY indicating their clinical relevance. However, future development of the progression statistic and confirmation of the genetic association in subjects from ongoing large studies such as ENROLL (27), with data collected more systematically than in REGISTRY but in less detail than TRACK-HD, would be ideal.

The genetic locus identified by the unified TRACK-HD progression measure association includes three genes, but *MSH3* is the likeliest candidate. Firstly, the lead SNP is a coding variant in exon 1 of

MSH3, MSH3 Pro67Ala, with the potential to affect function (SNiPA(28) accessed 10/11/2016). Clinically, each copy of the minor allele (G) at this SNP corresponds to a decrease of approximately 0.4 (95% CI=0.16,0.66) units per year in the rate of change of TMS, and a reduction of approximately 0.12 (95% CI=0.06,0.18) units per year in the rate of change of TFC (see Supplementary Information). Secondly, *MSH3* has been extensively implicated in the pathogenesis of HD in both mouse and cell studies, though this is the first human study to link *MSH3* to HD. MSH3 is a neuronally expressed member of a family of DNA mismatch repair proteins (29); it forms a heteromeric complex with MSH2 to form MutSβ, which recognises insertion-deletion loops of up to 13 nucleotides (30) **(Figure 4D)**. There is, however, a high level of interconnectedness between pathways involved in the DNA damage response, and MutSβ is implicated in other processes (20). Changes in CAG repeat size occur in terminally differentiated neurons in several HD mouse models and in human patient striatum, the brain area most affected in HD, and notably, somatic expansion of the CAG repeat in HD patient brain predicts onset (31). *Msh3* is required for both somatic expansion of *HTT* CAG repeats and for enhancing an early disease phenotype in mouse striatum (32), *Msh3* expression level is associated with repeat instability in mouse brain, (whereas DHFR is not) (30) and expansion of CAG and CTG repeats is prevented by msh3Δ in *Saccharomyces cerevisiae (33)*. This gives a plausible mechanism through which variation in *MSH3* could operate in HD **(Figure 4C and D)**. In patients with myotonic dystrophy type 1 (DM), somatic instability of the CTG repeat (CAG on the non-coding strand), is associated with age of onset and an *MSH3* variant was recently associated with somatic instability in blood DNA of patients (34). Variants in DNA repair pathways including those in *MSH3* contribute to age of onset modification of multiple CAG repeat expansion diseases (35) implicating the CAG repeat itself as the source of modification in these diseases.

This is the first study to use a measure of progression to look for modifiers of a neurodegenerative Mendelian disorder. We detected association with a coding variant on chromosome 5, reaching genome-wide significance given its annotation (21) in just 216 subjects, which replicated in a larger

independent sample and strengthened on meta-analysis. This indicates that either our progression measure developed in TRACK-HD is an excellent reflection of disease pathophysiological progression or that this is a locus with a very large effect size, or, most likely, both. While there are three genes at the locus, the most significant variant gives a coding change in MSH3, which together with the prior biological evidence makes it the most likely candidate. Somatic expansion of the CAG repeat through alterations in *MSH3* is a plausible mechanism for pathogenesis in HD which can be followed up in functional experiments in HD models. These data provide additional support for the therapeutic targeting of Huntingtin and the stability of its CAG repeat. Loss of or variation in mismatch repair complexes can cause malignancy and thus they are not regarded as ideal drug targets, but MSH3 is not essential as it can tolerate loss of function variation (36) and could provide a therapeutic target in HD. We note that if it does operate to alter repeat expansion it may also be a drug target in other repeat expansion disorders.

Acknowledgements and roles of funding sources

We would like to thank the people who have enabled this work through their participation in the TRACK-HD and REGISTRY studies.

We would like to thank the following organisations for their support of this project: The European Commission 7th Framework Program, (FP7/2007-2013) under grant agreement n° 2012-305121 "Integrated European –omics research project for diagnosis and therapy in rare neuromuscular and neurodegenerative diseases (NeurOmics)" who provided funding for this project. CHDI Foundation, Inc., a nonprofit biomedical research organization exclusively dedicated to developing therapeutics that will substantially improve the lives of HD-affected individuals who funded the TRACK-HD and REGISTRY studies. The Medical Research Council for their support of the MRC Centre for Neuropsychiatric Genetics and Genomics, MR/L010305/1. The Brain Research Trust (BRT), the Guarantors of Brain and the Medical Research Council UK who all supported this project.

The funders of the study and of the TRACK-HD and REGISTRY studies had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Author contributions and declarations

DJHM collected data, undertook analysis, and wrote the first draft of the ms. AFP undertook the genetic analysis, co-wrote the ms. DL undertook the statistical analysis of phenotype, co-wrote the ms. KL undertook genetic analysis. BRL collected data. RR collected data. AD collected data. SM co-supervised the genetic analysis. PH co-supervised data analyses, undertook genetic analysis, and co-wrote the ms. LJ helped secure funding, supervised data analyses, co-wrote the ms. SJT conceived the study, secured funding, recruited subjects, supervised data analyses and co-wrote the ms.

DL reports grant funding from CHDI via University College London (UCL), and personal fees from Roche Pharmaceutical, Voyager Pharmaceutical, and Teva Pharmaceuticals. BRL reports grants from CHDI Foundation via UCL, Teva Pharmaceuticals, and Lifemax Pharmaceuticals, and personal fees from Novartis, Roche, uniQure, Ionis Pharmaceuticals, and Raptor Pharmaceuticals. DJHM, KL, AD, AFP, SM, LJ, RR, PH, and SJT declare no competing interests.

Figure & Table legends

Figure 1: Study Design. After establishing that brain imaging, quantitative motor and cognitive variables are correlated and follow a similar trajectory, we scored the TRACK-HD subjects using principal component 1 as a Unified progression measure, and used this measure to look for genome-wide associations with HD progression. We replicated our findings in the EHDN Registry subjects by looking at how far their disease had progressed compared with expectations based on CAG/Age, and used this progression measure to look for genome-wide associations in REGISTRY. 1835

Registry subjects had genotype data (8). UHDRS TMS: Unified Huntington's Disease Rating Scale Total Motor Score. SDMT: symbol digit modality test. TFC: Total Functional Capacity.

Figure 2: Assessing progression in Huntington's disease (A) Graphical illustration of the trajectory of HD symptoms and signs over time, annotated to show what time period the different measures of onset and progression discussed in this paper cover. The TRACK-HD progression score uses longitudinal data over 3 years. Given limited longitudinal data in REGISTRY, cross-sectional severity at last visit compared to predicted severity was used as a proxy for progression. Age at onset occurs when a subject has unequivocal motor signs of Huntington's disease. (B) Distribution of progression measure in 218 members of TRACK-HD cohort. (C) Distribution of atypical severity (compared to predicted severity at final visit) in in 1835 members of the REGISTRY cohort. The curves in (B) and (C) are the normal distribution approximations of the severity score distributions.

Figure 3: Genome-wide Association Analysis of Progression Score. Green line in A-C: $5x10^{-8}$. (A) Manhattan plot of TRACK-HD GWA analysis yielding a locus on chromosome 5. Significance of SNPs (y axis) is plotted against genomic location (x axis). (B) Manhattan plot of REGISTRY GWA analysis showing suggestive trails on chromosome 15 in the same area as the GeM GWAS significant locus (8), and chromosome 5 in the same area as the TRACK progression GWAS. (C) Manhattan plot of Meta-analysis of TRACK and REGISTRY progression analysis. (D) Locus zoom plot of the TRACK-HD (top), REGISTRY (middle) and meta-analysis (bottom) data showing the structure of linkage disequilibrium (LD) and $-\log^{10}$(p-value) of the significant locus on chromosome. The top image shows the chromosome; the red square shows the region which is zoomed in on in the other panels. The colours of the circles are based on r^2 with the lead SNP in TRACK-HD as shown in the bottom of the plot; intensity of colour reflects multiple overlying SNPs. Dashed lines: $5x10^{-8}$

Figure 4: Significant genes are functionally linked and may cause somatic expansion of the *HTT* CAG repeat tract. STRING diagram showing all proteins from the Pearl *et al* (20) dataset with gene-wide p-values for association with Huntington's disease progression < 0.02 in **A**: the TRACK-HD dataset and **B**, the meta-analysis of TRACK-HD and REGISTRY (**http://hdresearch.ucl.ac.uk/data-resources/**). Genes with p<0.02 coloured; 10 further interactors in grey, confidence of interaction is shown in the 'Edge confidence' box, homo sapiens protein data used: http://string-db.org/cgi/ accessed October 2016 and January 2017 (37). **C** Schematic diagram showing how DNA mismatch repair proteins may be involved in somatic expansion of the CAG tract. Proteins with p<0.01 in the meta-analysed progression GWAS are coloured red. (i) The CAG repeat DNA is partly unwound by lesions, constraints of the CAG tract structure (middle image) or by transcription. (ii) This unwound DNA is recognised by MutSbeta (MSH2/MSH3) which recruits the endonuclease MutLalpha (PMS2/MLH1) and cleaves the DNA. (iii) Repair of the strand break leads to expansion of the CAG repeat. In neurones of the striatum somatic expansion is an ongoing process that occurs throughout life and variants in MSH3 may promote or inhibit repeat recognition, binding or repair. **D** Potential link between degree of somatic expansion over a patient's lifespan and rate of Huntington's disease progression.

Table 1: Setscreen enrichment p-values for the 14 pathways highlighted in GeM-HD (8).

The GO and KEGG terms in the first column refer to pathways of biologically related genes in the Gene Ontology Consortium(1) and Kyoto Encyclopedia of Genes and Genomes (2) databases respectively. The p-values in columns 2 – 4 refer to the association between the pathway indicated and rate of progression described in this paper (TRACK- TRACK-HD study; REGISTRY- REGISTRY study; META- meta-analysis). P(GeM) refers to the association between the indicated pathway and age at motor onset in the GeM-HD study (8).

Pathway	p(TRACK)	p(REGISTRY)	P(META)	p(GeM)	Description
GO: 32300	3·46E-09	8·34E-04	1.14E-11	3·82E-05	mismatch repair complex
KEGG 3430	2·79E-07	4·80E-02	1.34E-16	6·65E-06	mismatch repair (KEGG)
GO: 30983	6·66E-07	4·20E-04	3.17E-11	7·43E-06	mismatched DNA binding
GO: 6298	3·53E-06	4·59E-02	6.54E-09	3·25E-06	mismatch repair
GO: 32407	1·82E-02	1·10E-01	6.40E-04	5·74E-05	MutSalpha complex binding
GO: 32389	2·25E-02	4·69E-02	5.23E-04	1·66E-05	MutLalpha complex
GO: 33683	8·01E-02	5·87E-04	6.74E-03	1·69E-06	nucleotide-excision repair, DNA incision
GO: 90141	3·32E-01	5·93E-02	7.87E-01	2·30E-06	positive regulation of mitochondrial fission
GO: 1900063	4·10E-01	7·29E-01	6.93E-01	8·39E-05	regulation of peroxisome organization
GO: 90200	4·58E-01	5·44E-01	5.28E-01	8·89E-08	positive regulation of release of cytochrome c from mitochondria
GO: 90140	5·39E-01	3·32E-01	8.10E-01	1·57E-05	regulation of mitochondrial fission
GO: 10822	6·21E-01	6·28E-01	8.53E-01	7·63E-05	positive regulation of mitochondrion organization
GO: 4748	9·64E-01	6·97E-01	9.79E-01	2·66E-05	ribonucleoside-diphosphate reductase activity, thioredoxin disulfide as acceptor
GO: 16728	9·64E-01	6·97E-01	9.79E-01	2·66E-05	oxidoreductase activity, acting on CH or CH2 groups, disulfide as acceptor

REFERENCES

1. Huntington's, Disease, Collaborative, Research, Group. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group. Cell. 1993;72(6):971-83.
2. Ross CA, Tabrizi SJ. Huntington's disease: from molecular pathogenesis to clinical treatment. The Lancet Neurology. 2011;10(1):83-98.
3. Hogarth P, Kayson E, Kieburtz K, Marder K, Oakes D, Rosas D, et al. Interrater agreement in the assessment of motor manifestations of Huntington's disease. Movement disorders : official journal of the Movement Disorder Society. 2005;20(3):293-7.
4. Long JD, Paulsen JS, Marder K, Zhang Y, Kim JI, Mills JA. Tracking motor impairments in the progression of Huntington's disease. Movement disorders : official journal of the Movement Disorder Society. 2013;29(3):311-9.
5. Tabrizi SJ, Scahill RI, Owen G, Durr A, Leavitt BR, Roos RA, et al. Predictors of phenotypic progression and disease onset in premanifest and early-stage Huntington's disease in the TRACK-HD study: analysis of 36-month observational data. The Lancet Neurology. 2013;12(7):637-49.
6. Tabrizi SJ, Langbehn DR, Leavitt BR, Roos RA, Durr A, Craufurd D, et al. Biological and clinical manifestations of Huntington's disease in the longitudinal TRACK-HD study: cross-sectional analysis of baseline data. Lancet neurology. 2009;8(9):791-801.
7. Langbehn DR, Brinkman RR, Falush D, Paulsen JS, Hayden MR. A new model for prediction of the age of onset and penetrance for Huntington's disease based on CAG length. Clinical genetics. 2004;65(4):267-77.
8. Consortium GMoHsDG-H. Identification of Genetic Factors that Modify Clinical Onset of Huntington's Disease. Cell. 2015;162(3):516-26.
9. Orth M, Handley OJ, Schwenke C, Dunnett SB, Craufurd D, Ho AK, et al. Observing Huntington's Disease: the European Huntington's Disease Network's REGISTRY. PLoS currents. 2010;2:RRN1184.
10. Paulsen JS, Langbehn DR, Stout JC, Aylward E, Ross CA, Nance M, et al. Detection of Huntington's disease decades before diagnosis: the Predict-HD study. Journal of neurology, neurosurgery, and psychiatry. 2008;79(8):874-80.
11. Copyright (c) 2002-2012 by SAS Institute Inc. C, NC, USA. SAS. SAS Institute Inc. Cary, NC, USA.: SAS Institute Inc; 2002-2012.
12. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A Tool for Genome-wide Complex Trait Analysis. Am J Hum Genet. 2011;88(1):76-82.
13. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26(17):2190-1.
14. Pickrell JK, Berisa T, Liu JZ, Segurel L, Tung JY, Hinds DA. Detection and interpretation of shared genetic influences on 42 human traits. Nature genetics. 2016;48(7):709-17.
15. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS computational biology. 2015;11(4):e1004219.
16. Maston GA, Evans SK, Green MR. Transcriptional regulatory elements in the human genome. Annual review of genomics and human genetics. 2006;7:29-59.
17. Moskvina V, O'Dushlaine C, Purcell S, Craddock N, Holmans P, O'Donovan MC. Evaluation of an approximation method for assessment of overall significance of multiple-dependent tests in a genomewide association study. Genetic epidemiology. 2011;35(8):861-6.
18. Kiezun A, Garimella K, Do R, Stitziel NO, Neale BM, McLaren PJ, et al. Exome sequencing and the genetic basis of complex traits. Nature genetics. 2012;44(6):623-30.
19. Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science (New York, NY). 2015;348(6235):648-60.

20. Pearl LH, Schierz AC, Ward SE, Al-Lazikani B, Pearl FM. Therapeutic opportunities within the DNA damage response. Nat Rev Cancer. 2015;15(3):166-80.
21. Sveinbjornsson G, Albrechtsen A, Zink F, Gudjonsson SA, Oddson A, Masson G, et al. Weighting sequence variants based on their annotation increases power of whole-genome association studies. Nature genetics. 2016;48(3):314-7.
22. Plenge RM, Scolnick EM, Altshuler D. Validating therapeutic targets through human genetics. Nat Rev Drug Discov. 2013;12(8):581-94.
23. Trinh J, Gustavsson EK, Vilariño-Güell C, Bortnick S, Latourelle J, McKenzie MB, et al. DNM3 and genetic modifiers of age of onset in LRRK2 Gly2019Ser parkinsonism: a genome-wide linkage and association study. The Lancet Neurology. 2016;15(12):1248-56.
24. Corvol H, Blackman SM, Boelle PY, Gallins PJ, Pace RG, Stonebraker JR, et al. Genome-wide association meta-analysis identifies five modifier loci of lung disease severity in cystic fibrosis. Nature communications. 2015;6:8382.
25. Sham PC, Purcell SM. Statistical power and significance testing in large-scale genetic studies. Nature reviews Genetics. 2014;15(5):335-46.
26. Long JD, Langbehn DR, Tabrizi SJ, Landwehrmeyer BG, Paulsen JS, Warner J, et al. Validation of a prognostic index for Huntington's disease. Movement disorders : official journal of the Movement Disorder Society. 2017;32(2):256-63.
27. Landwehrmeyer GB, Fitzer-Attas CJ, Giuliano JD, Gonçalves N, Anderson KE, Cardoso F, et al. Data Analytics from Enroll-HD, a Global Clinical Research Platform for Huntington's Disease. Movement Disorders Clinical Practice. 2016.
28. Arnold M, Raffler J, Pfeufer A, Suhre K, Kastenmüller G. SNiPA: an interactive, genetic variant-centered annotation browser. Bioinformatics. 2015;31(8):1334-6.
29. Gonitel R, Moffitt H, Sathasivam K, Woodman B, Detloff PJ, Faull RL, et al. DNA instability in postmitotic neurons. Proceedings of the National Academy of Sciences of the United States of America. 2008;105(9):3467-72.
30. Tome S, Manley K, Simard JP, Clark GW, Slean MM, Swami M, et al. MSH3 Polymorphisms and Protein Levels Affect CAG Repeat Instability in Huntington's Disease Mice. Plos Genetics. 2013;9(2):16.
31. Swami M, Hendricks AE, Gillis T, Massood T, Mysore J, Myers RH, et al. Somatic expansion of the Huntington's disease CAG repeat in the brain is associated with an earlier age of disease onset. Hum Mol Genet. 2009;18(16):3039-47.
32. Dragileva E, Hendricks A, Teed A, Gillis T, Lopez ET, Friedberg EC, et al. Intergenerational and striatal CAG repeat instability in Huntington's disease knock-in mice involve different DNA repair genes. Neurobiol Dis. 2009;33(1):37-47.
33. Williams GM, Surtees JA. MSH3 Promotes Dynamic Behavior of Trinucleotide Repeat Tracts In Vivo. Genetics. 2015;200(3):737-+.
34. Morales F, Vasquez M, Santamaria C, Cuenca P, Corrales E, Monckton DG. A polymorphism in the MSH3 mismatch repair gene is associated with the levels of somatic instability of the expanded CTG repeat in the blood DNA of myotonic dystrophy type 1 patients. DNA repair. 2016;40:57-66.
35. Bettencourt C, Hensman-Moss D, Flower M, Wiethoff S, Brice A, Goizet C, et al. DNA repair pathways underlie a common genetic mechanism modulating onset in polyglutamine diseases. Ann Neurol. 2016;79(6):983-90.
36. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536(7616):285-91.
37. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic acids research. 2015;43(Database issue):D447-52.

Table 1: Setscreen enrichment p-values for the 14 pathways highlighted in GeM-HD (8). The GO and KEGG terms in the first column refer to pathways of biologically related genes in the Gene Ontology Consortium(1) and Kyoto Encyclopedia of Genes and Genomes (2) databases respectively. The p-values in columns 2 – 4 refer to the association between the pathway indicated and rate of progression described in this paper (TRACK- TRACK-HD study; REGISTRY- REGISTRY study; META- meta-analysis). P(GeM) refers to the association between the indicated pathway and age at motor onset in the GeM-HD study (8).

Pathway	p(TRACK)	p(REGISTRY)	P(META)	p(GeM)	Description
GO: 32300	3·46E-09	8·34E-04	1.14E-11	3·82E-05	mismatch repair complex
KEGG 3430	2·79E-07	4·80E-02	1.34E-16	6·65E-06	KEGG_MISMATCH_REPAIR
GO: 30983	6·66E-07	4·20E-04	3.17E-11	7·43E-06	mismatched DNA binding
GO: 6298	3·53E-06	4·59E-02	6.54E-09	3·25E-06	mismatch repair
GO: 32407	1·82E-02	1·10E-01	6.40E-04	5·74E-05	MutSalpha complex binding
GO: 32389	2·25E-02	4·69E-02	5.23E-04	1·66E-05	MutLalpha complex
GO: 33683	8·01E-02	5·87E-04	6.74E-03	1·69E-06	nucleotide-excision repair, DNA incision
GO: 90141	3·32E-01	5·93E-02	7.87E-01	2·30E-06	positive regulation of mitochondrial fission
GO: 1900063	4·10E-01	7·29E-01	6.93E-01	8·39E-05	regulation of peroxisome organization
GO: 90200	4·58E-01	5·44E-01	5.28E-01	8·89E-08	positive regulation of release of cytochrome c from mitochondria
GO: 90140	5·39E-01	3·32E-01	8.10E-01	1·57E-05	regulation of mitochondrial fission
GO: 10822	6·21E-01	6·28E-01	8.53E-01	7·63E-05	positive regulation of mitochondrion organization
GO: 4748	9·64E-01	6·97E-01	9.79E-01	2·66E-05	ribonucleoside-diphosphate reductase activity, thioredoxin disulfide as acceptor
GO: 16728	9·64E-01	6·97E-01	9.79E-01	2·66E-05	oxidoreductase activity, acting on CH or CH2 groups, disulfide as acceptor

1. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature genetics. 2000;25(1):25-9.
2. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research. 2000;28(1):27-30.

Identification of genetic variants associated with Huntington's disease progression: a genome-wide association study

Davina J Hensman Moss* , Antonio F. Pardiñas*, Douglas Langbehn, Kitty Lo, Blair R. Leavitt, Raymund Roos, Alexandra Durr, Simon Mead, Peter Holmans, Lesley Jones[§], Sarah J Tabrizi[§] and the REGISTRY and the TRACK-HD investigators. *Contributed equally; [§]Contributed equally.

The Lancet Neurology 2017

Supplementary Material: Table of Contents

Supplementary text information

Methods

Defining progression in TRACK-HD

Among the wide variety of potential cognitive and quantitative-motor variables, we analysed a subset of those that were previously used in a 36-month predefined primary analysis(1). A small number of quantitative-motor variables that were substantively redundant were eliminated and those with more tractable metric properties were chosen (**Supplementary Table 2**).

For the Track HD study, 10 subjects were excluded because they had no follow-up data. 15 other subjects were excluded because of missing brain MRI data there was no missing data for the other variables used in the analysis.

Our models controlled for study site, gender, education, and their interactions with follow-up time, consistent with the models used in the TRACK-HD standard analyses which are described elsewhere(1-4). The dominance of the first principal component is shown in the Scree plot in **Supplementary Figure 9**.

Progression analysis in REGISTRY

We used a square-root transform of TMS to improve approximate multivariate normality of the data. Missing data were considerable as documented **Supplementary Table 15**.

To deal with the missing data for clinical items, multiple imputation with 25 imputations was performed. Age, gender, and CAG expansion length were auxiliary variables for the imputations. Proper methods to account for imputation variation were used for all statistical inferences. Final parameter estimates and statistical significance were estimated by Rubin's method(5). We performed the above using the MI and MIANALYZE procedures of SAS/STAT 13.1(6).

In order to generate atypical severity scores, we needed to undertake three sequential procedures: (i) Multiple imputation of missing data (ii) Principal Component Analysis (PCA) and severity scoring of the combined imputed data replications (iii) Regression of the predictive effect of age, CAG length, and gender on the PCA-derived severity scores so that we are left with a measure of atypical (or “unexplained”) severity. The steps were taken in the order above; given that these steps could be done in different orders we also confirmed that there were only minimal differences due to the order (*data not shown*). We also noted some evidence of study site effects in the eventual regressions. Thus we used a random effect for site in models adjusting for age and CAG. Atypical severity was defined as the residual between each subject's observed and marginal predicted value. The dominance of the first principal component is shown in **Supplementary Figure 10**.

The final averaged multiple imputation model used a 2 degree of freedom restricted cubic spline(7) of cumulative probability of onset (CPO), plus main effects of gender and CAG length and a random effect for

site. Marginal effects from this model, which represent the estimated effects after accounting for site fluctuations, were used for all predictions. The knot placement for the clinical probability of onset spline was defined a priori using a conventional standard at the 10^{th}, 50^{th}, and 90^{th} percentiles of its observed distribution. The corresponding values were (0.131, 0.395, 0.885). Atypical severity was defined as the residual between each subject's observed and marginal predicted value. Final parameter estimates, along with estimates of statistical significance adjusted for the multiple imputation procedure are shown in the **Supplementary Table 16**.

We inspected the potential biasing influence of the CAG repeats, by classifying the individual in short (CAG < 41) and long (CAG > 55) repeats. We found an overrepresentation of people with larger atypical severity scores among those with short CAG, which implies that those with a small number of repeats are more likely to be in the study if atypically severely affected. This is likely to be due to the disease only being partially penetrant in those with short CAG repeats, resulting in bias (8). This prompted us to exclude subjects with short CAG from the creation of the severity scores, while retaining those with long CAG. However, we confirmed that the age-CAG severity function predicted using CAG > 41 gave sensible estimates for both the short and long ranges, enabling even those subjects with short CAG to be used in the final analysis (**Supplementary Figure 11)**.

Comparing TRACK-HD and REGISTRY progression measures

There are four common measures between TRACK-HD and REGISTRY: TMS, symbol digit score, Stroop word reading score and TFC. We took the first principal component score from an analysis of these four measures at the last TRACK-HD visit: this accounted for 79.4% of the variance in the PCA and correlated approximately equally with each of the four observed variables (**Supplementary Table 21**). To calculate the measure of severity unaccounted for by age and CAG length in TRACK, we regressed these principal component scores on the same predictors used for the unified REGISTRY progression measure, to give TRACK-HD severity scores.

As explained in the manuscript page 13, within the TRACK-HD data, the last-visit severity scores had a Pearson correlation of 0·674 with the previously calculated longitudinal progression measure. It can be shown that the predicted values obtained from the TRACK-HD and REGISTRY formulas are nearly linear, hence that Pearson correlation should be an adequate descriptive statistic for the relationship (**Supplementary Figure 12**).

Genotyping and quality control

DNA was obtained from blood samples of the 218 TRACK-HD study participants who had complete serial phenotype data, using standard methods (2). Genotyping was performed in Illumina Omni2.5 v1.1 arrays at

UCL Genomics, in accordance with the Infinium LCG Assay (15023141_A, June 2010) protocol (Illumina Inc, San Diego, USA). Standard QC procedures (9) were performed using PLINK v1.9 (10), including controlling for coverage and call rates (5% of missing data allowed per SNP and individual), inbreeding ($F < 0.2$ required) and Hardy-Weinberg equilibrium (SNPs with $p < 10^{-6}$ in an exact test were removed). With these criteria, and after removing one individual of a twin pair, a total of 216 gene positive TRACK-HD subjects were left in the sample, genotyped for 2.34 million genome-wide markers (**Figure 1**).

Identity-by-descent analysis showed 9 pairs of individuals with a relatedness coefficient ($\hat{\pi}$) higher than 0.15, which included 6 putative first degree relatives, 2 putative second degree relatives and 1 putative pair of third degree relatives. Additionally, an ADMIXTURE analysis with a subset of the 1000 Genomes (11) populations revealed 6 individuals with more than 25% of non-European ancestry. All these individuals were retained in the TRACK-HD sample, as their relatedness and admixture can be accommodated well by using association methods based on mixed linear models (12, 13).

TRACK-HD was imputed in the Cardiff University high-performance computing cluster RAVEN(14), using the SHAPEIT/IMPUTE2 algorithms(15, 16) and a standardised pipeline(17). The 1000 Genomes phase 3 panel provided by the IMPUTE2 authors (release October 2014), was used as the reference imputation panel. Imputation probabilities ("dosages") were converted to best-guess genotypes in fcGENE v1.07(18) using a minimum probability threshold of 80% and a per-SNP missingness threshold of 5% of the sample. After this process an INFO score cutoff of 0.8 was applied in order to select well-imputed variants, and all monomorphic and singleton markers were excluded. With these filters 9.65 million biallelic markers remained in the dataset.

Genotypes for the REGISTRY subjects were obtained from the GeM-HD Consortium (19), where details of their genotyping, curation and imputation are provided. This dataset harboured 8.94 million biallelic markers of 1,773 individuals (**Figure 1**).

Mixed linear model GWAS

Association analyses were performed with the mixed linear model (MLM) functions included in GCTA v1.26(20), specifically the leave-one-chromosome-out (LOCO) procedure(21). As the genetic relationship matrix used by MLMs can accurately account for cryptic relatedness and ancestry, and phenotypic variables already controlled for relevant clinical covariates, no covariates were added to the analyses. In order to transform the results into independent GWAS signals, PLINK was again used to perform linkage disequilibrium (LD) clumping ($r^2 = 0.1$, $p < 1x10^{-4}$; window size < 3 Mb). Due to the relatively small size of the TRACK-HD and REGISTRY samples, calculation of SNP-based heritability (h^2_{SNP}) for our tested phenotypes was not possible using either genotyped or imputed markers(22, 23). Because of the small sample sizes, analyses were restricted to SNPs with minor allele frequency >1%.

Meta-analysis of the GWAS summary statistics from the TRACK-HD and REGISTRY studies was carried out using the fixed effects method with inverse-variance weights as implemented in METAL (24). The meta-analysis of TRACK-HD and REGISTRY studies was carried out using the fixed effects method with inverse-variance weights as implemented in METAL(24). To control for spurious results due to scale differences between the TRACK-HD and REGISTRY progression phenotypes, effect sizes from both summary statistics were standardised to have equal variances before meta-analysis.

QQ plots of observed log p-values (sorted by value) for each SNP versus their expected values in the absence of association are shown for TRACK-HD, REGISTRY and the meta-analysis in **Supplementary Figure 13**. If there is no association, and no systematic inflation in the test statistics (for example, from population stratification), the observed log p-values would follow their expected values (the red line in **Supplementary Figure 13**) exactly. Indeed, this is what is observed for the majority of data points, which do not show association. The extent to which such systematic inflation exists is measured by the genomic

95% confidence interval for log p-values in the absence of association is shaded grey, and the points lying above this in the top right corner indicate genuine associations.

Conditional analyses of GWAS summary statistics were carried out using the COJO procedure included in GCTA v1.26(26).

Co-localisation analyses

In order to discern if our top GWAS signals were mediated by the same SNPs in both TRACK-HD and REGISTRY, we used the co-localisation method of Giambartolomei *et al*.(27), as implemented in GWAS-pw v0.21 (28). In summary, the GWAS summary statistics of our two samples were first divided into approximately independent LD blocks(29), and each block was then scanned to estimate the probability (in a hierarchical Bayesian framework) of harbouring an association common to the two samples. In contrast to the original algorithm, the model priors do not need to be pre-specified in GWAS-pw, as they are estimated directly from the summary statistics. This implementation has been thoroughly tested by simulation and applied to real data from heterogeneous sources (28). By testing the entire genome instead of a small number of candidate regions arising from the GWAS clumps, we follow a conservative approach towards estimating co-localisation, which also has the desirable property of allowing us to compare our candidates (to the resolution of single SNPs) with every other region in the genome.

A similar procedure was used to test for co-localisation between the region on chromosome 5 containing GWAS signal in TRACK-HD and REGISTRY and SNPs influencing expression (eQTLs), since this may indicate which gene in an association region is causal. Given that eQTLs close to the gene (cis-eQTLs) tend to replicate more reliably than those from other parts of the genome (30), these analyses were restricted to

the regions of GWAS signal and genes within 1Mb of these regions. These analyses used expression data from 53 tissues, accessed through GTeX (31). To minimise multiple testing, the two tissues showing the most significant eQTLs for each gene were used for the co-localisation analysis. Additionally, for DHFR and MSH3, analyses were performed using three brain tissues (caudate, cerebellum and cortex), since these are the most biologically relevant to HD a priori. Co-localisation results are shown for the TRACK-HD GWAS in **Supplementary Table 8**, and the REGISTRY GWAS in **Supplementary Table 9**. Plots of GWAS and eQTL signals with significant co-localisation are shown in in **Supplementary Figures 7 and 8.**

Gene-based and gene-set analyses

Gene-wide p-values were calculated using MAGMA v1.05 (32) on the TRACK-HD and REGISTRY summary statistics, by summing the p-values of all SNPs inside each gene. MAGMA aggregates the association evidence across all SNPs in a gene, while correcting for LD between SNPs (using the European data from Phase 3 of the 1000 Genomes Project as reference). This analysis increases power when a gene contains multiple causal SNPs (e.g. as a result of allelic heterogeneity), or when the causal SNP is not typed and its signal is partially captured by multiple genotyped SNPs in LD with it. We set a window of 35 kb upstream and 10 kb downstream of each gene in order to capture the signal of proximal regulatory SNPs(33, 34).

To maximise comparability with the GeM GWAS, our primary gene-set analyses used Setscreen (Moskvina et al. 2011). Setscreen sums the (log-) p-values of all SNPs in the gene set, similar to Fisher's method, but adjusts the distribution to allow for non-independence of SNPs due to linkage disequilibrium (Brown 1975). Significant enrichments from the Setscreen analyses were confirmed using the competitive gene-set analysis procedure implemented in MAGMA. This more conservative approach tests whether genes in a gene set have more significant gene-wide p-values than other genes, correcting for gene size, SNP density and intergenic linkage disequilibrium (de Leeuw et al. 2015), but may be less powerful than the Setscreen analysis for small gene sets.

Initially, we performed gene set analyses on the 14 pathways found to be significantly enriched for association signal in the GeM GWAS. Many of these pathways relate to DNA repair, so we investigated the biological specificity of this signal further by analysing 78 gene-sets taken from a recent review of DNA repair (Pearl et al 2015).

As a secondary analysis, to potentially uncover areas of novel disease-related biology, we tested the same gene sets used by GeM-HD Consortium (2015). This comprises a collection of 14,706 pathways containing between 3 and 500 genes from the Gene Ontology (GO)(35), Kyoto Encyclopedia of Genes and Genomes (KEGG)(36), Mouse Genome Informatics (MGI)(37), National Cancer Institute (NCI)(38), Protein ANalysis THrough Evolutionary Relationships (PANTHER)(39), BioCarta(40) and Reactome(41). Multiple testing correction was carried out for this analysis by calculating q-values (Storey and Tibshirani, 2003).

Linking genetic variation to clinical measures

To explain how our TRACK-HD lead variant (rs557874766) affected commonly used clinical measures of HD severity we first correlated TRACK-HD progression score with UHDRS Total Motor Score (TMS) and UHDRS Total Functional Capacity (TFC). We defined "raw" TMS rate as TMS change divided by follow-up years and "adjusted" TMS rate as the residual of raw TMS rate after regressing off effects of initial TMS, age, sex, CAG. We followed the same procedure for TFC.

Regressing these measures on progression gives the following estimates of the amount of change for one unit increase in progression (standard errors in brackets):

Raw TMS rate: 0.71(0.19)
Adjusted TMS rate: 0.57 (0.18)
Raw TFC rate: 0.21 (0.047)
Adjusted TFC rate: 0.20 (0.044)

The effect size at the top MSH3 SNP in TRACK (rs557874766) is -0.58 (s.e. =0.087) units of progression per copy of the minor allele G (see **Supplementary Table 21**) – this corresponds to a change of -0.33 (95% CI =0.10, 0.56) to -0.41 (0.16,0.66) units in TMS rate compared to the major allele C, which can be interpreted as a reduction in the rate of TMS increase by 0.33-0.41 units per year for each copy of the G allele. Similarly, this corresponds to a reduction in the rate of TFC change of 0.12 (0.06,0.18) units per year per G allele.

Results

Since *MSH3* is a member of all the most significantly enriched pathways, we tested whether *MSH3* was individually responsible for the pathway enrichments by removing it and repeating the analyses. GO:32300 and KEGG:3430 are still nominally significant in TRACK (p=0.0413, p=0.0452 respectively) but not in REGISTRY. Neither of the two Pearl pathways is significant in TRACK or REGISTRY. The only pathways nominally significant both in TRACK and REGISTRY are GO:32389 (MutLalpha complex) and Pearl pathway "Repair_pathway/SSR/MMR/MutL_homologs", neither of which contain *MSH3*. Thus, it appears that the mismatch repair pathway enrichments are mainly driven by *MSH3*. However, in the TRACK-REGISTRY meta-analysis, the Pearl et al. MMR pathway ($p=1.27\times10^{-4}$), GO:32300 ($p=1.02\times10^{-3}$), KEGG 3430 (1.07×10^{-4}) and GO:30983 are at least nominally significant without *MSH3*. Pathway enrichments without *MSH3* are shown in **Supplementary Table 18** for the 14 GeM pathways and **Supplementary Table 19** for the Pearl et al. pathways.

Setscreen gene set analysis of the large set of pathways analysed by the GeM-HD Consortium (2015) is shown in **Supplementary Table 24.** There were 26 pathways showing significant ($q<0.05$) enrichment in

22. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. Nat Genet. 2010;42(7):565-9.
23. Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AAE, Lee SH, et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat Genet. 2015;47(10):1114-20.
24. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26(17):2190-1.
25. Devlin B, Roeder K. Genomic control for association studies. Biometrics. 1999;55(4):997-1004.
26. Yang J, Ferreira T, Morris AP, Medland SE, Madden PA, Heath AC, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat Genet. 2012;44(4):369-75.
27. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. PLoS Genet. 2014;10(5):e1004383.
28. Pickrell JK, Berisa T, Liu JZ, Segurel L, Tung JY, Hinds DA. Detection and interpretation of shared genetic influences on 42 human traits. Nat Genet. 2016;48(7):709-17.
29. Berisa T, Pickrell JK. Approximately independent linkage disequilibrium blocks in human populations. Bioinformatics. 2016;32(2):283-5.
30. Ramasamy A, Trabzuni D, Guelfi S, Varghese V, Smith C, Walker R, et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. Nature neuroscience. 2014;17(10):1418-28.
31. Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science (New York, NY). 2015;348(6235):648-60.
32. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS computational biology. 2015;11(4):e1004219.
33. Maston GA, Evans SK, Green MR. Transcriptional Regulatory Elements in the Human Genome. Annu Rev Genomics Hum Genet. 2006;7(1):29-59.
34. The Network and Pathway Analysis Subgroup of the Psychiatric Genomics Consortium. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. Nat Neurosci. 2015;18(2):199-209.
35. Consortium TGO. Gene Ontology Consortium: going forward. Nucleic Acids Res. 2015;43(D1):D1049-D56.
36. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 2016;44(D1):D457-D62.
37. Eppig JT, Blake JA, Bult CJ, Kadin JA, Richardson JE, Group TMGD. The Mouse Genome Database (MGD): facilitating mouse as a model for human biology and disease. Nucleic Acids Res. 2015;43(D1):D726-D36.
38. Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M, Hannay T, et al. PID: the Pathway Interaction Database. Nucleic Acids Res. 2009;37(Database issue):D674-D9.
39. Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the PANTHER classification system. Nat Protocols. 2013;8(8):1551-66.
40. Nishimura D. BioCarta. Biotech Software & Internet Report. 2001;2(3):117-20.
41. Fabregat A, Sidiropoulos K, Garapati P, Gillespie M, Hausmann K, Haw R, et al. The Reactome pathway Knowledgebase. Nucleic Acids Res. 2016;44(Database issue):D481-D7.

The TRACK-HD Investigators

AUSTRALIA
Monash University, Victoria: J Stout, C Pourchot, J C Campbell, S C Andrews, A O'Regan, I Labuschagne, C Milchman, M Campbell, S Queller, E Frajman.

CANADA
University of British Columbia, Vancouver: A Coleman, R Dar Santos, J Decolongon, A Sturrock

FRANCE
APHP, Hôpital Salpêtriere, Paris: E Bardinet, C Jauffret, D Justo, S Lehericy, C Marelli, K Nigaud, R Valabrègue

GERMANY

TRACK after correction for multiple testing of pathways. These pathways mainly relate to DNA repair and binding, and none is more significant than GO:32300 (mismatch repair complex). The genes in these 26 pathways are shown in **Supplementary Table 25,** and are similar to those in Tables 2 and 3, with the exception of DHFR (however, the pathways containing DHFR tend to be less strongly associated than the mismatch repair pathways in both TRACK and REGISTRY). Thus, analysis of the large set of pathways does not appear to throw up any novel areas of biology outside those indicated by the GeM paper.

References

1. Tabrizi SJ, Scahill RI, Owen G, Durr A, Leavitt BR, Roos RA, et al. Predictors of phenotypic progression and disease onset in premanifest and early-stage Huntington's disease in the TRACK-HD study: analysis of 36-month observational data. The Lancet Neurology. 2013;12(7):637-49.
2. Tabrizi SJ, Langbehn DR, Leavitt BR, Roos RA, Durr A, Craufurd D, et al. Biological and clinical manifestations of Huntington's disease in the longitudinal TRACK-HD study: cross-sectional analysis of baseline data. Lancet neurology. 2009;8(9):791-801.
3. Tabrizi SJ, Scahill RI, Durr A, Roos RA, Leavitt BR, Jones R, et al. Biological and clinical changes in premanifest and early stage Huntington's disease in the TRACK-HD study: the 12-month longitudinal analysis. Lancet neurology. 2011;10(1):31-42.
4. Tabrizi SJ, Reilmann R, Roos RA, Durr A, Leavitt B, Owen G, et al. Potential endpoints for clinical trials in premanifest and early Huntington's disease in the TRACK-HD study: analysis of 24 month observational data. Lancet neurology. 2012;11(1):42-53.
5. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York: Wiley; 2008.
6. Inc. SI. SAS/STAT 13.1 User's Guide: High-Performance Procedures. . In: Inc. SI, editor. NC2013.
7. Harrell FE. Regression Modeling Strategies. New York: Wiley; 2001.
8. Langbehn DR, Brinkman RR, Falush D, Paulsen JS, Hayden MR. A new model for prediction of the age of onset and penetrance for Huntington's disease based on CAG length. Clinical genetics. 2004;65(4):267-77.
9. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. Nat Protocols. 2010;5(9):1564-73.
10. Chang CC, Chow CC, Tellier L, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4(7).
11. 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491(7422):56-65.
12. Thornton T, Conomos MP, Sverdlov S, Blue EM, Cheung CY, Glazner CG, et al. Estimating and adjusting for ancestry admixture in statistical methods for relatedness inference, heritability estimation, and association testing. BMC Proceedings. 2014;8(1):1-7.
13. Shin J, Lee C. A mixed model reduces spurious genetic associations produced by population stratification in genome-wide association studies. Genomics. 2015;105(4):191-6.
14. Advanced Research Computing @ Cardiff (ARCCA). Introduction to RAVEN: accessed: 29/03/2016; [Available from: http://www.cardiff.ac.uk/arcca/services/equipment/ravenintroduction.html.
15. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet. 2012;44(8):955-9.
16. Delaneau O, Zagury J-F, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. Nature Methods. 2013;10(1):5-6.
17. van Leeuwen EM, Kanterakis A, Deelen P, Kattenberg MV, The Genome of the Netherlands C, Slagboom PE, et al. Population-specific genotype imputations using minimac or IMPUTE2. Nat Protocols. 2015;10(9):1285-96.
18. Roshyara NR, Scholz M. fcGENE: a versatile tool for processing and transforming SNP datasets. PloS one. 2014;9(7):e97589.
19. Consortium GMoHsDG-H. Identification of Genetic Factors that Modify Clinical Onset of Huntington's Disease. Cell. 2015;162(3):516-26.
20. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A Tool for Genome-wide Complex Trait Analysis. Am J Hum Genet. 2011;88(1):76-82.
21. Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL. Advantages and pitfalls in the application of mixed-model association methods. Nat Genet. 2014;46(2):100-6.

University of Münster, Münster: N Bechtel, S Bohlen, R Reilmann;
University of Bochum, Bochum: A Hoffman, P Kraus;
University of Ulm, Ulm: B Landwehrmeyer.

NETHERLANDS
Leiden University Medical Centre, Leiden: SJA van den Bogaard, E M Dumas, J van der Grond, EP t'Hart, C Jurgens, M-N Witjes-Ane.

U.K.
St Mary's Hospital, Manchester: N Arran, J Callaghan, D Craufurd, C Stopford;
London School of Hygiene and Tropical Medicine, London: C Frost, R Jones;
University College London, London: N Hobbs, N Lahiri, R Ordidge, G Owen, T Pepple, J Read, M J Say, E Wild, A Patel, N C Fox, C Gibbard, I Malone, H Crawford, D Whitehead; R Scahill;
Imperial College London, London: S Keenan;
IXICO, London: D M Cash;
University of Oxford, Oxford: C Berna, S Hicks, C Kennard.

U.S.A.
University of Iowa, Iowa City, IA: E Axelson, H Johnson, C Wang, T Acharya;
Massachusetts General Hospital, Harvard, MA: S Lee, W Monaco, HD Rosas;
Indiana University, IN: C Campbell, S Queller, K Whitlock;
CHDI Foundation, New York, NY: B Borowsky; AJ Tobin.

The REGISTRY Investigators 2004-2012

Registry Steering committee: Anne-Catherine Bachoud-Lévi, Anna Rita Bentivoglio, Ida Biunno, Raphael Bonelli, Jean-Marc Burgunder, Stephen Dunnett, Joaquim Ferreira, Olivia Handley, Arvid Heiberg, Torsten Illmann, G. Bernhard Landwehrmeyer, Jamie Levey, Maria A. Ramos-Arroyo, Jørgen Nielsen, Susana Pro Koivisto, Markku Päivärinta, Raymund A.C. Roos, A Rojo Sebastián, Sarah Tabrizi, Wim Vandenberghe, Christine Verellen-Dumoulin, Tereza Uhrova, Jan Wahlström+, Jacek Zaremba

Language coordinators: Verena Baake, Katrin Barth, Monica Bascuñana Garde, Sabrina Betz, Reineke Bos, Jenny Callaghan, Adrien Come, Leonor Correia Guedes, Daniel Ecker, Ana Maria Finisterra, Ruth Fullam, Mette Gilling, Lena Gustafsson, Olivia J Handley, Carina Hvalstedt, Christine Held, Kerstin Koppers, Claudia Lamanna, Matilde Laurà, Asunción Martínez Descals, Saül Martinez-Horta, Tiago Mestre, Sara Minster, Daniela Monza, Lisanne Mütze, Martin Oehmen, Michael Orth, Hélène Padieu, Laurent Paterski, Nadia Peppa, Susana Pro Koivisto, Martina Di Renzo, Amandine Rialland, Niini Røren, Pavla Šašinková, Erika Timewell, Jenny Townhill, Patricia Trigo Cubillo, Wildson Vieira da Silva, Marleen R van Walsem, Carina Whalstedt, Marie-Noelle Witjes-Ané, Grzegorz Witkowski , Abigail Wright, Daniel Zielonka, Eugeniusz Zielonka, Paola Zinzi

AUSTRIA

Graz (Medizinische Universitäts Graz, Psychiatrie): Raphael M. Bonelli, Sabine Lilek, Karen Hecht, Brigitte Herranhof, Anna Holl (formerly Hödl), Hans-Peter Kapfhammer, Michael Koppitz, Markus Magnet, Nicole Müller, Daniela Otti, Annamaria Painold, Karin Reisinger, Monika Scheibl, Helmut Schöggl, Jasmin Ullah

Innsbruck (Universitätsklinik Innsbruck, Neurologie): Eva-Maria Braunwarth, Florian Brugger, Lisa Buratti, Eva-Maria Hametner, Caroline Hepperger, Christiane Holas, Anna Hotter, Anna Hussl, Christoph Müller, Werner Poewe, Klaus Seppi, Fabienne Sprenger,
Gregor Wenning

BELGIUM

Bierbeek: Andrea Boogaerts, Godelinde Calmeyn, Isabelle Delvaux, Dirk Liessens, Nele Somers

Charleroi (Institut de Pathologie et de Génétique (IPG)): Michel Dupuit, Cécile Minet, Dominique van Paemel, Pascale Ribaï, Christine Verellen-Dumoulin

Leuven: (Universitair Ziekenhuis Gasthuisberg,): Andrea Boogaerts, Wim Vandenberghe, Dimphna van Reijen

CZECH REPUBLIC

Prague (Extrapyramidové centrum, Neurologická klinika, 1. LF UK a VFN):
Jan Roth, Irena Stárková

DENMARK

Copenhagen (Neurogenetics Clinic, Danish Dementia Research Centre, Rigshospitalet, University of Copenhagen): Lena E. Hjermind, Oda Jacobsen, Jørgen E. Nielsen, Ida Unmack Larsen, Tua Vinther-Jensen

FINLAND

Turku-Suvituuli (Rehabilitation Centre Suvituuli): Heli Hiivola, Hannele Hyppönen, Kirsti Martikainen, Katri Tuuha

FRANCE

Angers (Centre de référence des maladies neurogénétique- CHU d'Angers): Philippe Allain, Dominique Bonneau, Marie Bost, Bénédicte Gohier, Marie-Anne Guérid, Audrey Olivier, Adriana Prundean, Clarisse Scherer-Gagou, Christophe Verny

Bordeaux (Hôpital CHU Pellegrin): Blandine Babiloni, Sabrina Debruxelles, Charlotte Duché, Cyril Goizet, Laetitia Jameau, Danielle Lafoucrière, Umberto Spampinato

Lille-Amiens :

Lille (CHRU Roger Salengro) : Rekha Barthélémy, Christelle De Bruycker, Maryline Cabaret, Anne-Sophie Carette, Eric Decorte Luc Defebvre, Marie Delliaux, Arnaud Delval, Alain Destee, Kathy Dujardin, Marie-Hélène Lemaire, Sylvie Manouvrier, Mireille Peter, Lucie Plomhouse, Bernard Sablonnière, Clémence Simonin, Stéphanie Thibault-Tanchou, Isabelle Vuillaume

Amiens (CHU Nord) : Marcellin Bellonet, Hassan Berrissoul, Stéphanie Blin, Françoise Courtin, Cécile Duru, Véronique Fasquel, Olivier Godefroy, Pierre Krystkowiak, Béatrice Mantaux, Martine Roussel, Sandrine Wannepain

Marseille (Hôpital La Timone) : Jean-Philippe Azulay, Marie Delfini, Alexandre Eusebio, Frédérique Fluchere, Laura Mundler

Strasbourg (Hôpital Civil) : Mathieu Anheim, Celine Julié, Ouhaid Lagha Boukbiza, Nadine Longato, Gabrielle Rudolf, Christine Tranchant, Marie-Agathe Zimmermann

GERMANY

Aachen (Universitätsklinikum Aachen, Neurologische Klinik): Christoph Michael Kosinski, Eva Milkereit, Daniela Probst, Kathrin Reetz, Christian Sass, Johannes Schiefer, Christiane Schlangen, Cornelius J. Werner

Berlin (Klinik und Poliklinik für Neurologie - Charité - Universitätsmedizin Berlin): Harald Gelderblom, Josef Priller, Harald Prüß, Eike Jakob Spruth

Bochum (Huntington-Zentrum (NRW) Bochum im St. Josef-Hospital): Gisa Ellrichmann, Lennard Herrmann, Rainer Hoffmann, Barbara Kaminski, Peter Kotz, Christian Prehn, Carsten Saft

Dinslaken (Reha Zentrum in Dinslaken im Gesundheitszentrums Lang): Herwig Lange, Robert Maiwald

Dresden (Universitätsklinikum Carl Gustav Carus an der Technischen Universität Dresden, Klinik und Poliklinik für Neurologie): Matthias Löhle, Antonia Maass, Simone Schmidt, Cecile Bosredon, Alexander Storch, Annett Wolz, Martin Wolz

Freiburg (Universitätsklinik Freiburg, Neurologie): Philipp Capetian, Johann Lambeck, Birgit Zucker

Hamburg (Universitätsklinikum Hamburg-Eppendorf, Klinik und Poliklinik für Neurologie): Kai Boelmans, Christos Ganos, Walburgis Heinicke, Ute Hidding, Jan Lewerenz, Alexander Münchau, Michael Orth, Jenny Schmalfeld, Lars Stubbe, Simone Zittel

Hannover (Neurologische Klinik mit Klinischer Neurophysiologie, Medizinische Hochschule Hannover): Gabriele Diercks, Dirk Dressler, Heike Gorzolla, Christoph Schrader, Pawel Tacik

Itzehoe (Schwerpunktpraxis Huntington, Neurologie und Psychiatrie): Michael Ribbat

Marburg KPP (Klinik für Psychiatrie und Psychotherapie Marburg-Süd): Bernhard Longinus

Marburg Uni (Universität Marburg, Neurologie): Katrin Bürk, Jens Carsten Möller, Ida Rissling

München (Huntington-Ambulanz im Neuro-Kopfzentrum - Klinikum rechts der Isar der Neurologischen Klinik und Poliklinik der Technischen Universität München): Mark Mühlau, Alexander Peinemann, Michael Städtler, Adolf Weindl, Juliane Winkelmann, Cornelia Ziegler

Münster (Universitätsklinikum Münster, Klinik und Poliklinik für Neurologie): Natalie Bechtel, Heike Beckmann, Stefan Bohlen, Eva Hölzner, Herwig Lange, Ralf Reilmann, Stefanie Rohm, Silke Rumpf , Sigrun Schepers, Natalia Weber

Taufkirchen (Isar-Amper-Klinikum - Klinik Taufkirchen (Vils)): Matthias Dose, Gabriele Leythäuser, Ralf Marquard, Tina Raab, Alexandra Wiedemann

Ulm (Universitätsklinikum Ulm, Neurologie): Katrin Barth, Andrea Buck, Julia Connemann, Daniel Ecker, Carolin Geitner, Christine Held, Andrea Kesse, Bernhard Landwehrmeyer, Christina Lang, Jan Lewerenz, Franziska Lezius, Solveig Nepper, Anke Niess, Michael Orth, Ariane Schneider, Daniela Schwenk, Sigurd Süßmuth, Sonja Trautmann, Patrick Weydt

ITALY

Bari Clinica Neurologica - Neurophysiopatology of Pain Unit UNIVERSITA' DI BARI): Claudia Cormio, Vittorio Sciruicchio, Claudia Serpino, Marina de Tommaso

Bologna (DIBINEM - Alma Mater Studiorum - Università di Bologna; IRCCS Istituto delle Scienze Neurologiche di Bologna): Sabina Capellari, Pietro Cortelli, Roberto Galassi, Rizzo Giovanni, Roberto Poda, Cesa Scaglione

Florence (Dipartimento di Scienze Neurologiche e Psichiatriche Universita' degli Studi di Firenze-Azienda Ospedaliera Universitaria Careggi): Elisabetta Bertini, Elena Ghelli, Andrea Ginestroni, Francesca Massaro, Claudia Mechi, Marco Paganini, Silvia Piacentini, Silvia Pradella, Anna Maria Romoli, Sandro Sorbi

Genoa (Dipartimento di Neuroscienze, Riabilitazione, Oftalmologia, Genetica e Scienze Materno-Infantili, Università di Genova): Giovanni Abbruzzese, Monica Bandettini di Poggio, Giovanna Ferrandes, Paola Mandich, Roberta Marchese

Milan (Fondazione IRCCS Istituto Neurologico Carlo Besta):
Alberto Albanese, Daniela Di Bella, Anna Castaldo, Stefano Di Donato, Cinzia Gellera, Silvia Genitrini, Caterina Mariotti, Daniela Monza, Lorenzo Nanetti, Dominga Paridi, Paola Soliveri, Chiara Tomasello

Naples (Dipartimento di Neuroscienze, Scienze Riproduttive e Odontostomatologiche, Università Federico II): Giuseppe De Michele, Luigi Di Maio, Marco Massarelli, Silvio Peluso, Alessandro Roca, Cinzia Valeria Russo, Elena Salvatore, Pierpaolo Sorrentino

Pozzilli (IS) (Centro di Neurogenetica e Malattie Rare - IRCCS Neuromed): Enrico Amico, Mariagrazia Favellato, Annamaria Griguoli, Irene Mazzante, Martina Petrollini, Ferdinando Squitieri and **Rome (Lega Italiana Ricerca Huntington e malattie correlate - onlus / www.LIRH.it):** Barbara D'Alessio, Chiara Esposito

Rome (Istituto di Farmacologia Traslazionale & Istituto di Scienze e Tecnologie della Cognizione /CNR, Istituto di Neurologia Università Cattolica del Sacro Cuore): Anna Rita Bentivoglio, Marina Frontali, Arianna Guidubaldi, Tamara Ialongo, Gioia Jacopini, Carla Piano, Silvia Romano, Francesco Soleti, Maria Spadaro, Paola Zinzi

NETHERLANDS

Enschede (Medisch Spectrum Twente): Monique S.E. van Hout, Marloes E. Verhoeven, Jeroen P.P. van Vugt, A. Marit de Weert

Groningen (Polikliniek Neurologie): J.J.W. Bolwijn, M. Dekker, B. Kremer, K.L. Leenders, J.C.H. van Oostrom

Leiden (Leiden University Medical Centre (LUMC)): Simon J. A. van den Bogaard, Reineke Bos, Eve M. Dumas, Ellen P. 't Hart, Raymund A.C. Roos

Nijmegen (Universitair Medisch Centrum St. Radboud, Neurology): Berry Kremer, C.C.P. Verstappen

NORWAY

Oslo University Hospital (Rikshospitalet, Dept. of Medical Genetics and Dept. of Neurology): Olaf Aaserud, Jan Frich C., Arvid Heiberg, Marleen R. van Walsem, Ragnhild Wehus

Oslo University Hospital (Ulleval, **Dept. of Medical Genetics and Dept.of Neurorehabilitation)**: Kathrine Bjørgo, Madeleine Fannemel, Per F. Gørvell, Eirin Lorentzen, Susana Pro Koivisto, Lars Retterstøl, Bodil Stokke

Trondheim (St. Olavs Hospital): Inga Bjørnevoll, Sigrid Botne Sando

POLAND

Gdansk (St. Adalbert Hospital, Gdansk, Medical University of Gdansk, Neurological and Psychiatric Nursing Dpt.): Artur Dziadkiewicz, Malgorzata Nowak, Piotr Robowski, Emilia Sitek, Jaroslaw Slawek, Witold Soltan, Michal Szinwelski

Katowice (Medical University of Silesia, Katowice): Magdalena Blaszcyk, Magdalena Boczarska-Jedynak, Ewelina

Daniel Stompel

Krakow (Krakowska Akademia Neurologii):

Wasielewska, Magdalena Wójcik

Poznan (Poznan University of Medical Sciences, Poland): Anna Bryl, Anna Ciesielska, Aneta Klimberg, Jerzy

Warsaw-MU (Medical University of Warsaw, Neurology): Anna Gogol (formerly Kalbarczyk), Piotr Janik, Hubert Kwiecinski, Zygmunt Jamrozik

Warsaw-IPiN (Institute of Psychiatry and Neurology Dep. of Genetics, First Dep. of Neurology): Jakub Antczak, Katarzyna Jachinska, Wioletta Krysa, Maryla Rakowicz, Przemyslaw Richter, Rafal Rola, Danuta Ryglewicz, Halina

Zieora-Jakutowicz

PORTUGAL

Coimbra (Hospital Universitário de Coimbra): Cristina Januário, Filipa Júlio

Lisbon (Clinical Pharmacology Unit, Instituto de Medicina Molecular, Faculty of Medicine, University of Lisbon): Joaquim J Ferreira, Miguel Coelho, Leonor Correia Guedes, Tiago Mendes, Tiago Mestre, Anabela Valadas

Porto (Hospital de São João, (Faculdade de Medicina da Universidade do Porto)): Carlos Andrade, Miguel Gago, Carolina Garrett, Maria Rosália Guerra.

SPAIN

Badajoz (Hospital Infanta Cristina): Carmen Durán Herrera, Patrocinio Moreno Garcia

Barcelona-Hospital Mútua de Terrassa : Miquel Aguilar Barbera, Dolors Badenes Guia, Laura Casas Hernanz , Judit López Catena, Pilar Quiléz Ferrer, Ana Rojo Sebastián, Gemma Tome Carruesco

Barcelona-Bellvitge (Hospital Universitari de Bellvitge): Jordi Bas, Núria Busquets, Matilde Calopa

Barcelona-Merced (Hospital Mare de Deu de La Merced): Misericordia Floriach Robert, Celia Mareca Viladrich, Jesús Miguel Ruiz Idiago, Antonio Villa Riballo

Burgos (Servicio de Neurología Hospital General Yagüe): Esther Cubo, Cecilia Gil Polo, Natividad Mariscal Perez, Jessica Rivadeneyra

Granada (Hospital Universitario San Cecilio, Neurología): Francisco Barrero, Blas Morales

Madrid-Clinico (Hospital Clínico Universitario San Carlos): María Fenollar, Rocío García-Ramos García, Paloma Ortega, Clara Villanueva

Madrid RYC (Hospital Ramón y Cajal, Neurología): Javier Alegre, Mónica Bascuñana, Juan Garcia Caldentey, Marta Fatás Ventura, Guillermo García Ribas, Justo García de Yébenes, José Luis López-Sendón Moreno, Patricia Trigo Cubillo

Madrid FJD (Madrid-Fundación Jiménez Díaz): Javier Alegre, Fernando Alonso Frech, Justo García de Yébenes, Pedro J García Ruíz, Asunción Martínez-Descals, Rosa Guerrero, María José Saiz Artiga, Vicenta Sánchez

Murcia (Hospital Universitario Virgen de la Arrixaca): María Fuensanta Noguera Perea, Lorenza Fortuna, Salvadora Manzanares, Gema Reinante, María Martirio Antequera Torres, Laura Vivancos Moreau

Oviedo (Hospital Central de Asturias): Sonia González González, Luis Menéndez Guisasola, Carlos Salvador, Esther Suaréz San Martín

Palma de Mallorca (Hospital Universitario Son Espases): Inés Legarda Ramirez, Aranzazú Gorospe, Mónica Rodriguez Lopera, Penelope Navas Arques, María José Torres Rodríguez, Barbara Vives Pastor

Pamplona (Complejo Hospitalario de Navarra): Itziar Gaston, Maria Dolores Martinez-Jaurrieta, Maria A. Ramos-Arroyo

Sevilla ("Hospital Virgen Macarena"): Jose Manuel Garcia Moreno, Carolina Mendez Lucena, Fatima Damas Hermoso, Eva Pacheco Cortegana, José Chacón Peña, Luis Redondo

Sevilla (Hospital Universitario Virgen del Rocío): Fátima Carrillo, María Teresa Cáceres, Pablo Mir, María José Lama Suarez, Laura Vargas-González

Valencia (Hospital la Fe): Maria E. Bosca, Francisco Castera Brugada, Juan Andres Burguera, Anabel Campos Garcia, Carmen Peiró Vilaplana

SWEDEN

Göteborg (Sahlgrenska University Hospital): Peter Berglund, Radu Constantinescu, Gunnel Fredlund, Ulrika Høsterey-Ugander, Petra Linnsand, Liselotte Neleborn-Lingefjärd, Jan Wahlström+, Magnus Wentzel

Umeå (Umeå University Hospital): Ghada Loutfi, Carina Olofsson, Eva-Lena Stattin, Laila Westman, Birgitta Wikström

SWITZERLAND

Bern: Jean-Marc Burgunder, Yanik Stebler **(Swiss HD Zentrum)**, Alain Kaelin, Irene Romero, Michael Schüpbach, Sabine Weber Zaugg **(Zentrum für Bewegungsstörungen, Neurologische Klinik und Poliklinik, Universität Bern)**

Zürich (Department of Neurology, University Hospital Zürich): Maria Hauer, Roman Gonzenbach, Hans H. Jung, Violeta Mihaylova, Jens Petersen

U.K.

Aberdeen (NHS Grampian Clinical Genetics Centre & University of Aberdeen): Roisin Jack, Kirsty Matheson, Zosia Miedzybrodzka, Daniela Rae, Sheila A Simpson, Fiona Summers, Alexandra Ure, Vivien Vaughan

Birmingham (The Barberry Centre, Dept of Psychiatry): Shahbana Akhtar, Jenny Crooks, Adrienne Curtis, Jenny de Souza (Keylock), John Piedad, Hugh Rickards, Jan Wright

Bristol (North Bristol NHs Trust, Southmead hospital): Elizabeth Coulthard, Louise Gethin, Beverley Hayward, Kasia Sieradzan, Abigail Wright

Cambridge (Cambridge Centre for Brain Repair, Forvie Site): Matthew Armstrong, Roger A. Barker, Deidre O'Keefe, Anna Di Pietro, Kate Fisher, Anna Goodman, Susan Hill, Ann Kershaw, Sarah Mason, Nicole Paterson, Lucy Raymond, Rachel Swain, Natalie Valle Guzman

Cardiff (Schools of Medicine and Biosciences, Cardiff University): Monica Busse, Cynthia Butcher, Jenny Callaghan, Stephen Dunnett, Catherine Clenaghan, Ruth Fullam, Olivia Handley, Sarah Hunt, Lesley Jones, Una Jones, Hanan Khalil, Sara Minster, Michael Owen, Kathleen Price, Anne Rosser, Jenny Townhill

Edinburgh (Molecular Medicine Centre, Western General Hospital, Department of Clinical Genetics): Maureen Edwards, Carrie Ho (Scottish Huntington´s Association), Teresa Hughes (Scottish Huntington´s Association), Marie McGill, Pauline Pearson, Mary Porteous, Paul Smith (Scottish Huntington´s Association)

Fife (Scottish Huntington's Association Whyteman's Brae Hospital): Peter Brockie, Jillian Foster, Nicola Johns, Sue McKenzie, Jean Rothery, Gareth Thomas, Shona Yates

Gloucester (Department of Neurology Gloucestershire Royal Hospital): Liz Burrows, Carol Chu, Amy Fletcher, Deena Gallantrae, Stephanie Hamer, Alison Harding, Stefan Klöppel, Alison Kraus, Fiona Laver, Monica Lewis, Mandy Longthorpe, Ivana Markova, Ashok Raman, Nicola Robertson, Mark Silva, Aileen Thomson, Sue Wild, Pam Yardumian

Hull (Castle Hill Hospital): Carol Chu, Carole Evans, Deena Gallentrae, Stephanie Hamer, Alison Kraus, Ivana Markova, Ashok Raman

Leeds (Chapel Allerton Hospital, Department of Clinical Genetics): Leeds (Chapel Allerton Hospital, Clinical Genetics): Carol Chu, Stephanie Hamer, Emma Hobson, Stuart Jamieson, Alison Kraus, Ivana Markova, Ashok Raman, Hannah Musgrave, Liz Rowett, Jean Toscano, Sue Wild, Pam Yardumian

Leicester (Leicestershire Partnership Trust, Mill Lodge): Colin Bourne, Jackie Clapton, Carole Clayton, Heather Dipple, Dawn Freire-Patino, Janet Grant, Diana Gross, Caroline Hallam, Julia Middleton, Ann Murch, Catherine Thompson

Liverpool (Walton Centre for Neurology and Neurosurgery): Sundus Alusi, Rhys Davies, Kevin Foy, Emily Gerrans, Louise Pate

London (Guy's Hospital): Thomasin Andrews, Andrew Dougherty, Charlotte Golding, Fred Kavalier, Hana Laing, Alison Lashwood, Dene Robertson, Deborah Ruddy, Alastair Santhouse, Anna Whaite

London (The National Hospital for Neurology and Neurosurgery): Thomasin Andrews, Stefania Bruno, Karen Doherty, Charlotte Golding, Salman Haider, Davina Hensman, Nayana Lahiri, Monica Lewis, Marianne Novak, Aakta Patel, Nicola Robertson, Elisabeth Rosser, Sarah Tabrizi, Rachel Taylor, Thomas Warner, Edward Wild

Manchester (Genetic Medicine, University of Manchester, Manchester Academic Health Sciences Centre and Central Manchester University Hospitals NHS Foundation Trust): Natalie Arran, Judith Bek, Jenny Callaghan, David Craufurd, Ruth Fullam, Marianne Hare, Liz Howard, Susan Huson, Liz Johnson, Mary Jones, Helen Murphy, Emma Oughton, Lucy Partington-Jones, Dawn Rogers, Andrea Sollom, Julie Snowden, Cheryl Stopford, Jennifer Thompson, Iris Trender-Gerhard, Nichola Verstraelen (formerly Ritchie), Leann Westmoreland

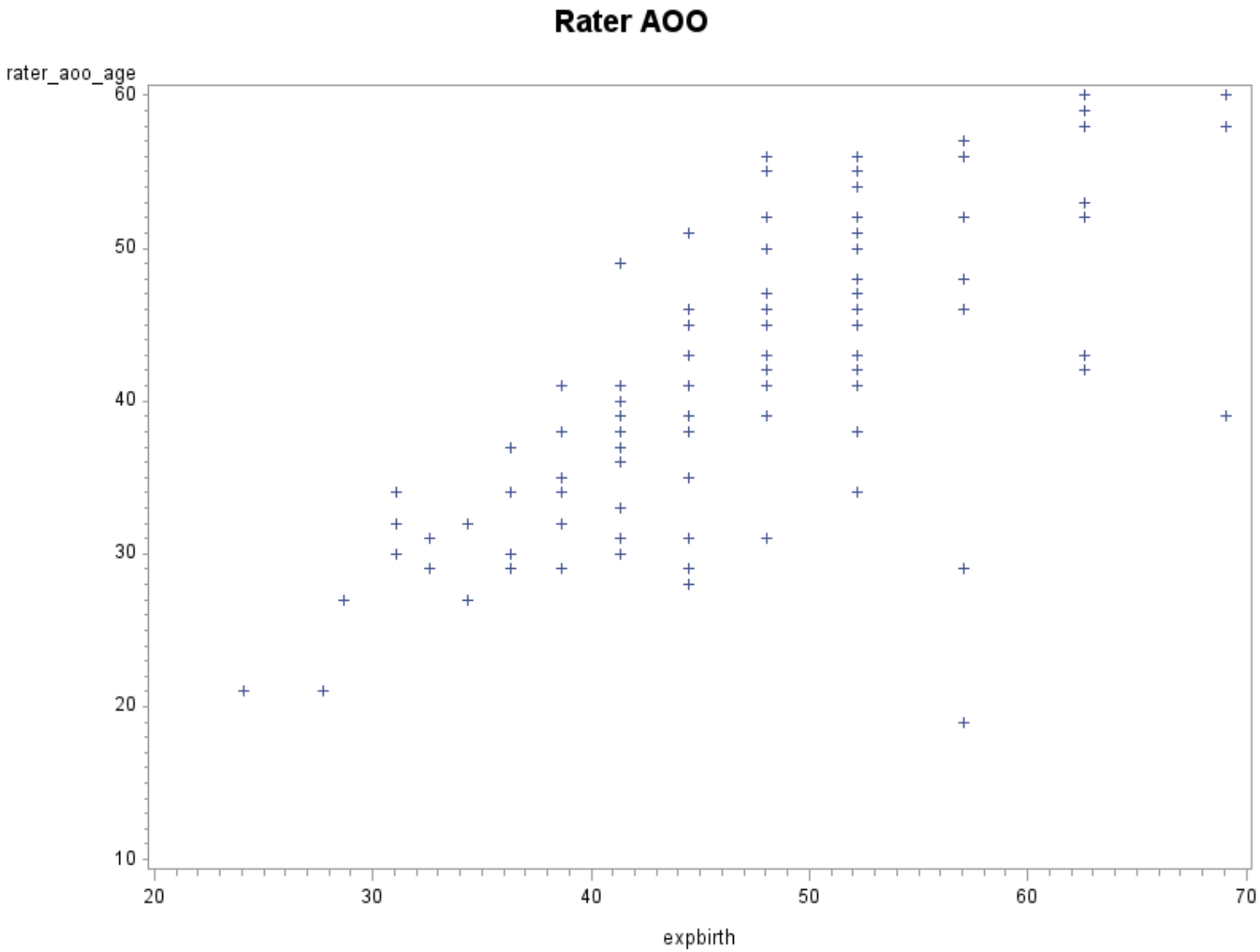
Oxford (Oxford University Hospitals NHS Trust, Dept. of Neurosciences, University of Oxford): Richard Armstrong, Kathryn Dixon, Andrea H Nemeth, Gill Siuda, Ruth Valentine

Plymouth (Plymouth Huntington Disease Service, Mount Gould Hospital): David Harrison, Max Hughes, Andrew Parkinson, Beverley Soltysiak

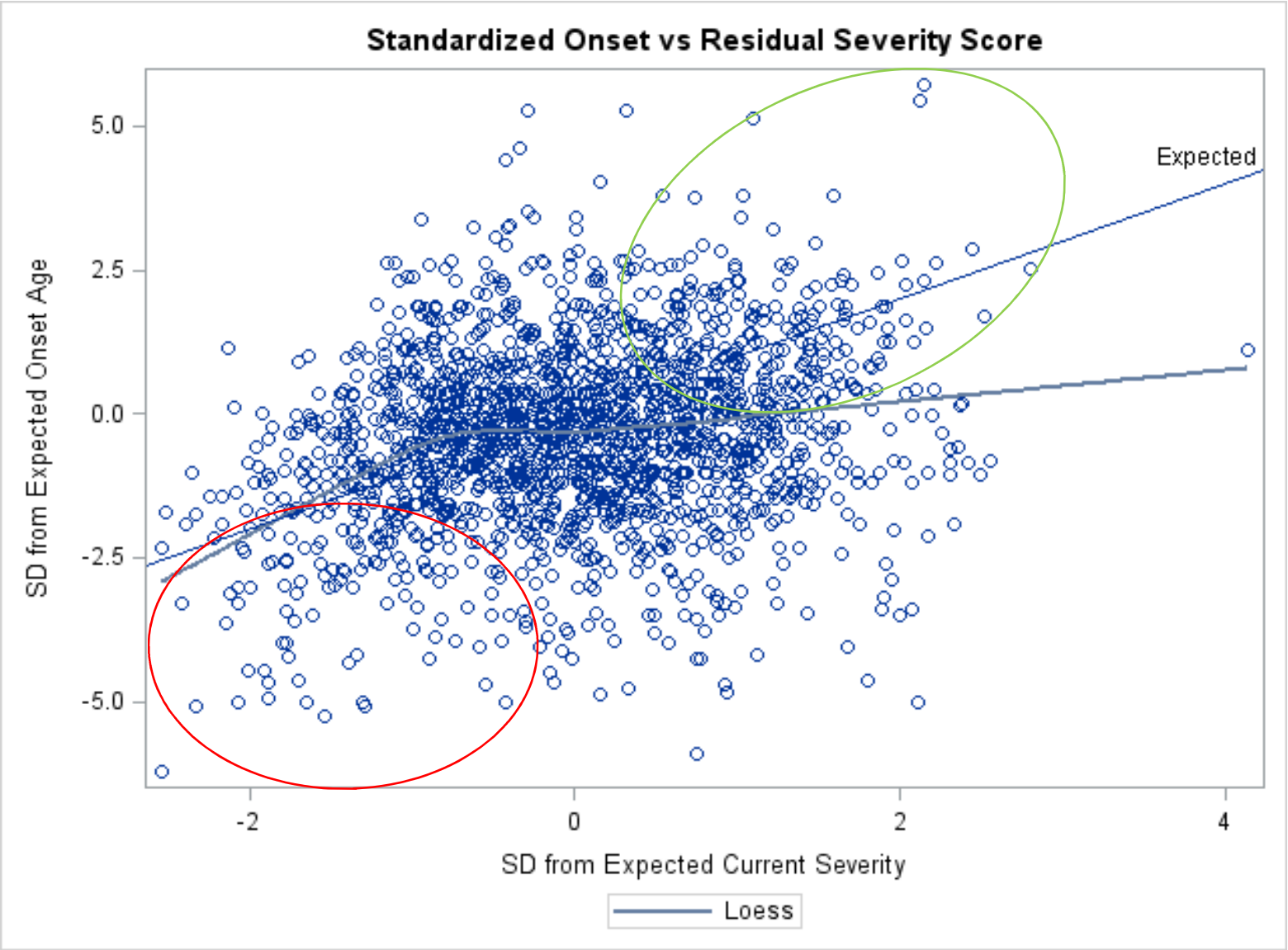
Sheffield (The Royal Hallamshire Hospital– Sheffield Children's Hospital): Oliver Bandmann, Alyson Bradbury, Paul Gill, Helen Fairtlough, Kay Fillingham, Isabella Foustanos, Mbombe Kazoka, Kirsty O'Donovan, Nadia Peppa, Cat Taylor, Katherine Tidswell, Oliver Quarrell

EHDN's associate site in SINGAPORE: National Neuroscience Institute Singapore: Jean-Marc Burgunder, Puay Ngoh Lau, Emmanul Pica, Louis Tan

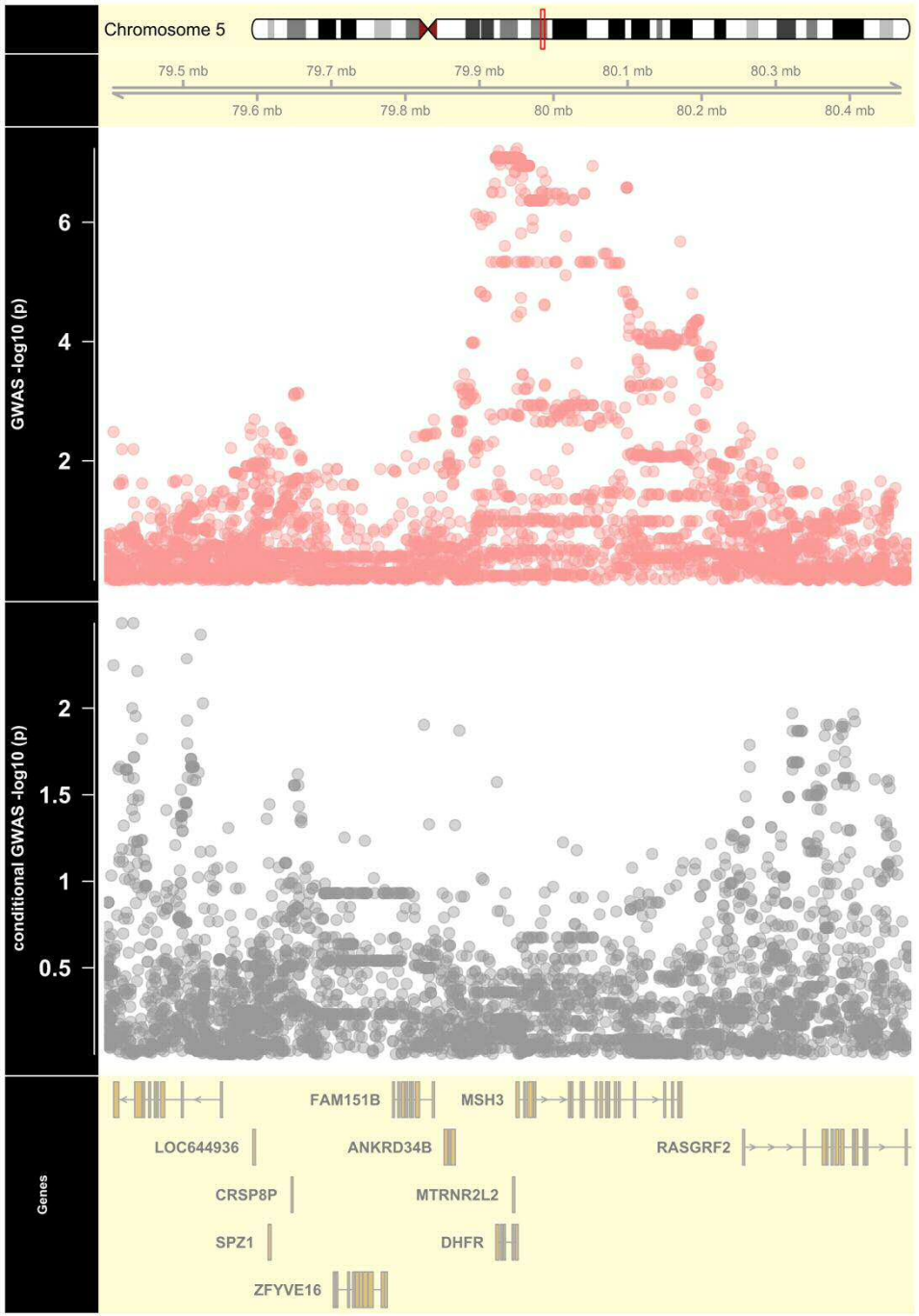
Supplementary Figure 1: Observed versus Expected Age of Onset Among Those Who Have Experienced Onset in the TRACK-HD analysis: amongst these 96 subjects who had experienced onset, the rater AAO showed the expected relation with predicted AAO based on CAG length. Earlier than predicted onset age was correlated with faster progression (using the unified HD progression measure) (r=-0·315; p = 0·002)



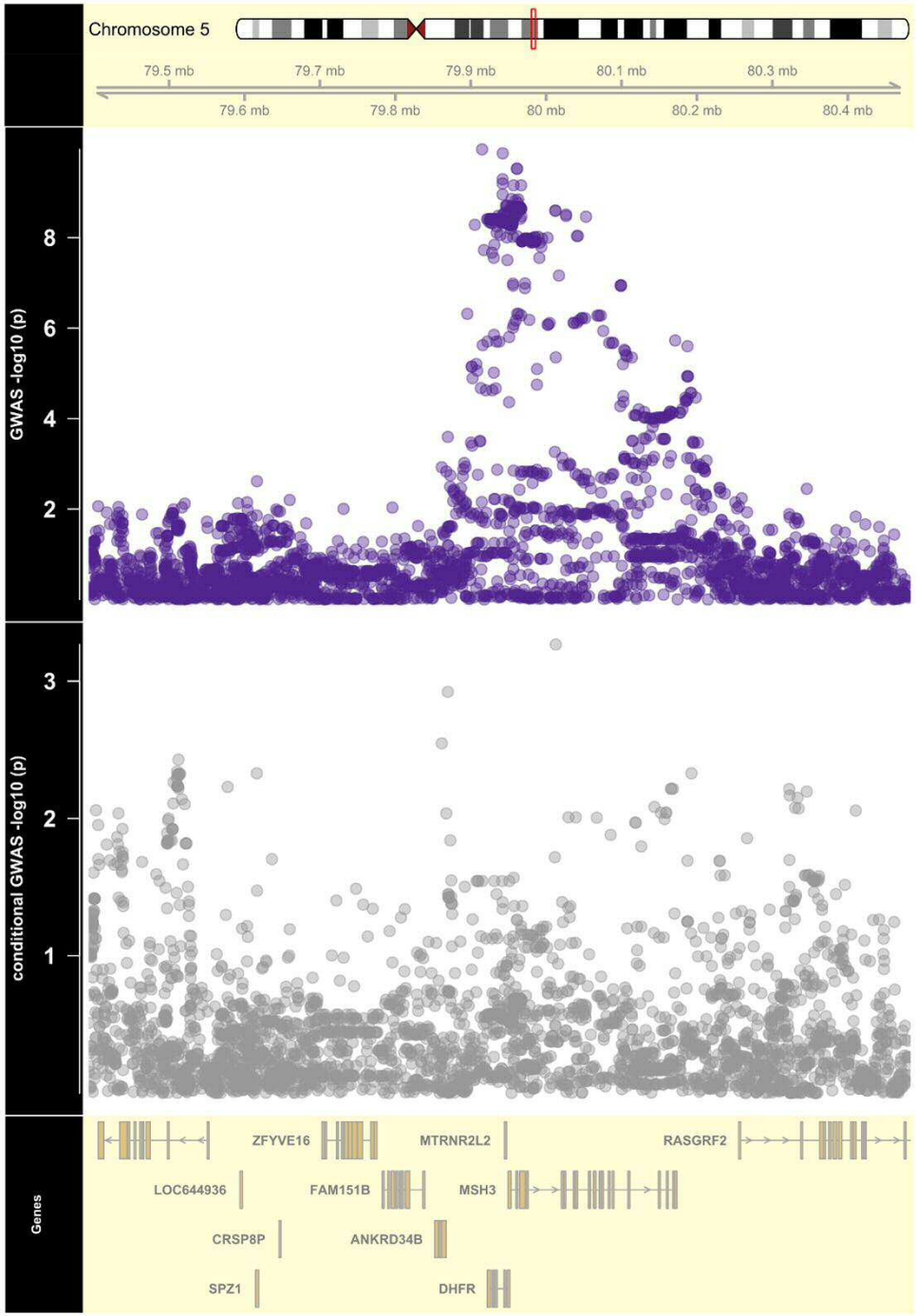
Supplementary Figure 2: REGISTRY progression measure and atypical onset age are modestly correlated in REGISTRY. Note bias for very late expected onset for those with low CAG repeats. SD = Standard deviation.



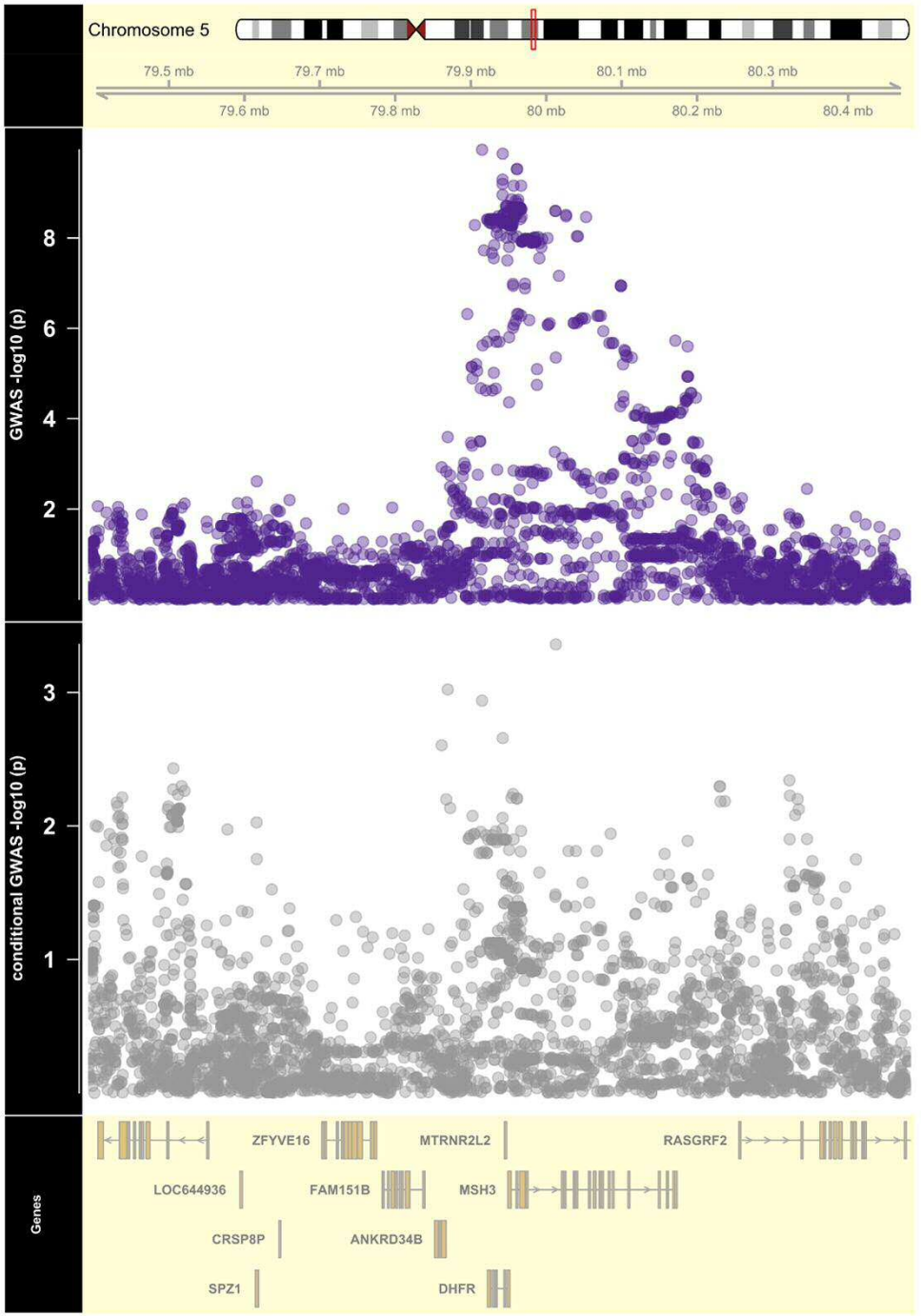
Supplementary Figure 3: Regional plot of TRACK-HD GWAS signal in the MSH3-DHFR region before(top) and after (bottom) conditioning on the most significant SNP in TRACK-HD (rs557874766). The lack of significant association after conditioning on this SNP is consistent with here being only one association signal in the region.



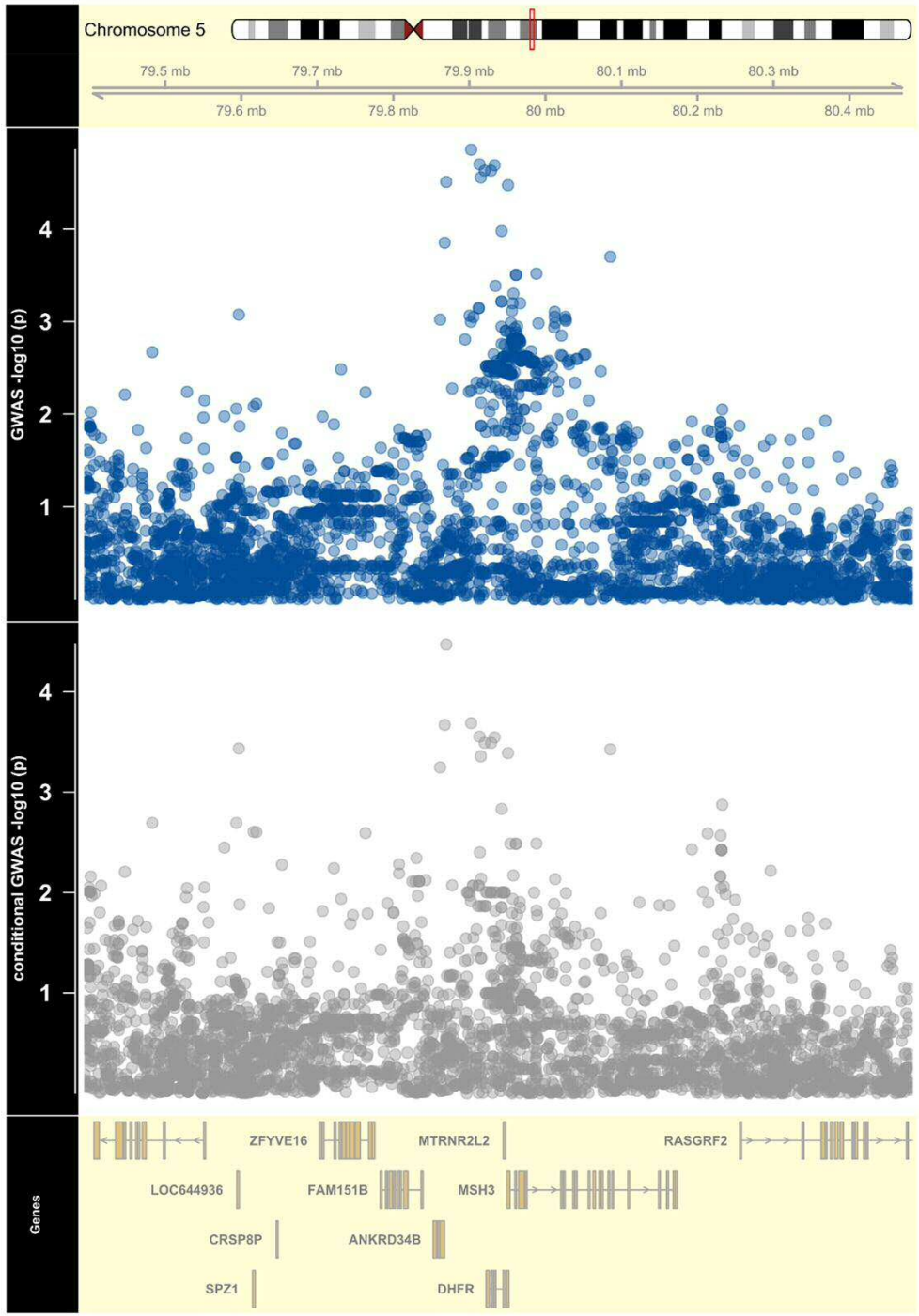
Supplementary Figure 4: Regional plot of TRACK-HD and REGISTRY meta-analysis GWAS signal in the MSH3-DHFR region before(top) and after (bottom) conditioning on the most significant SNP in the meta-analysis (rs1232027). The lack of significant association after conditioning on this SNP is consistent with here being only one association signal in the region.



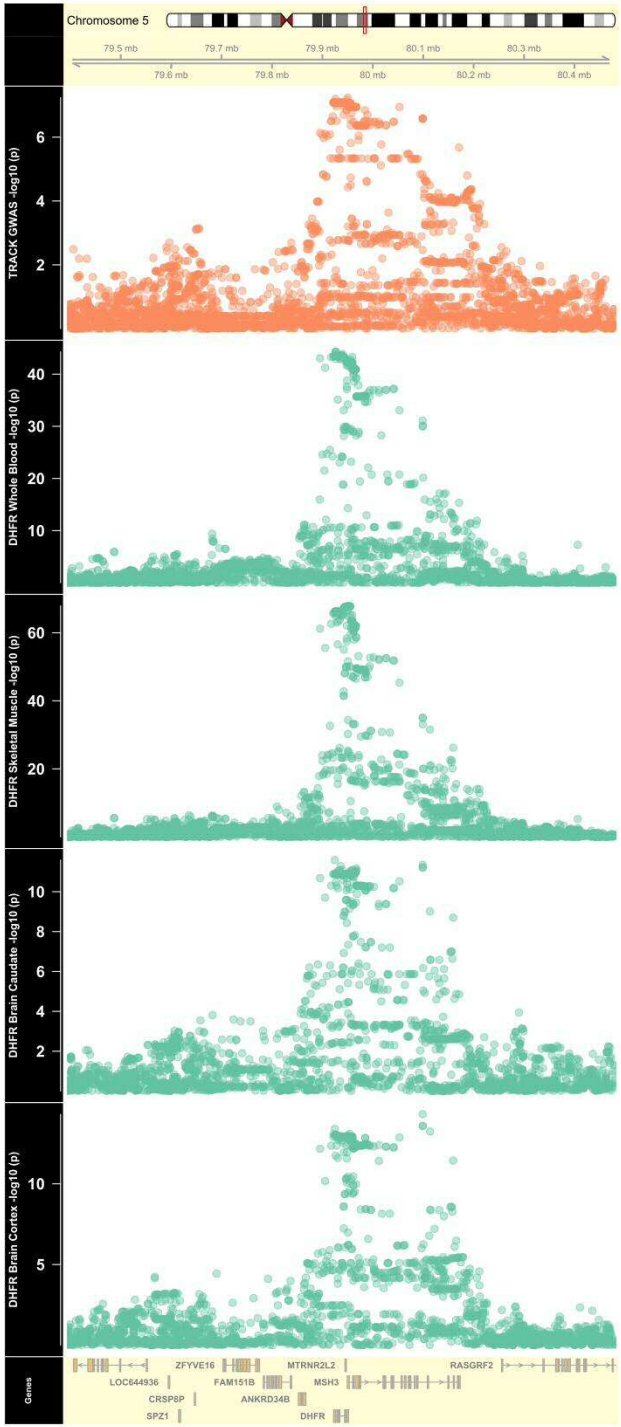
Supplementary Figure 5: Regional plot of TRACK-HD and REGISTRY meta-analysis GWAS signal in the MSH3-DHFR region before(top) and after (bottom) conditioning on the most significant SNP in TRACK-HD (rs557874766). The lack of significant association after conditioning on this SNP is consistent with here being only one association signal in the region.



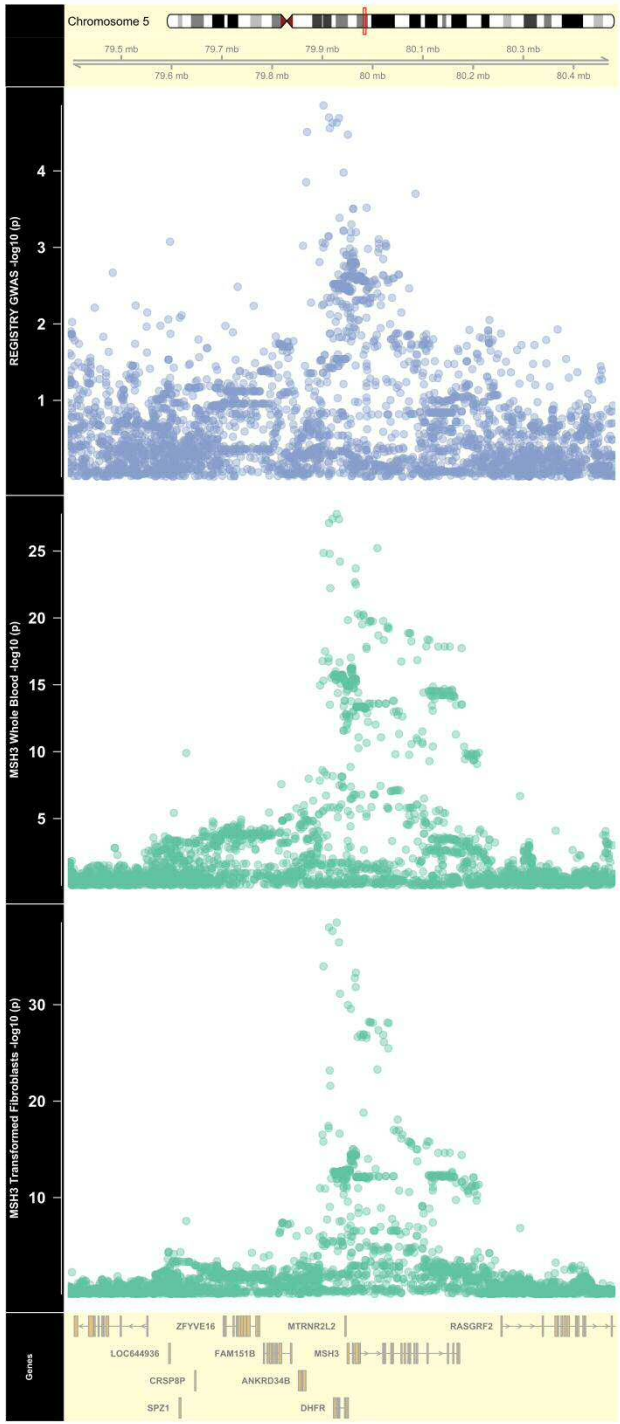
Supplementary Figure 6: Regional plot of REGISTRY GWAS signal in the MSH3-DHFR region before(top) and after (bottom) conditioning on the most significant SNP in TRACK-HD (rs557874766). The significance of association is largely unaffected by conditioning on this SNP. This indicates that rs557874766 does not explain the REGISTRY association signal in this region.



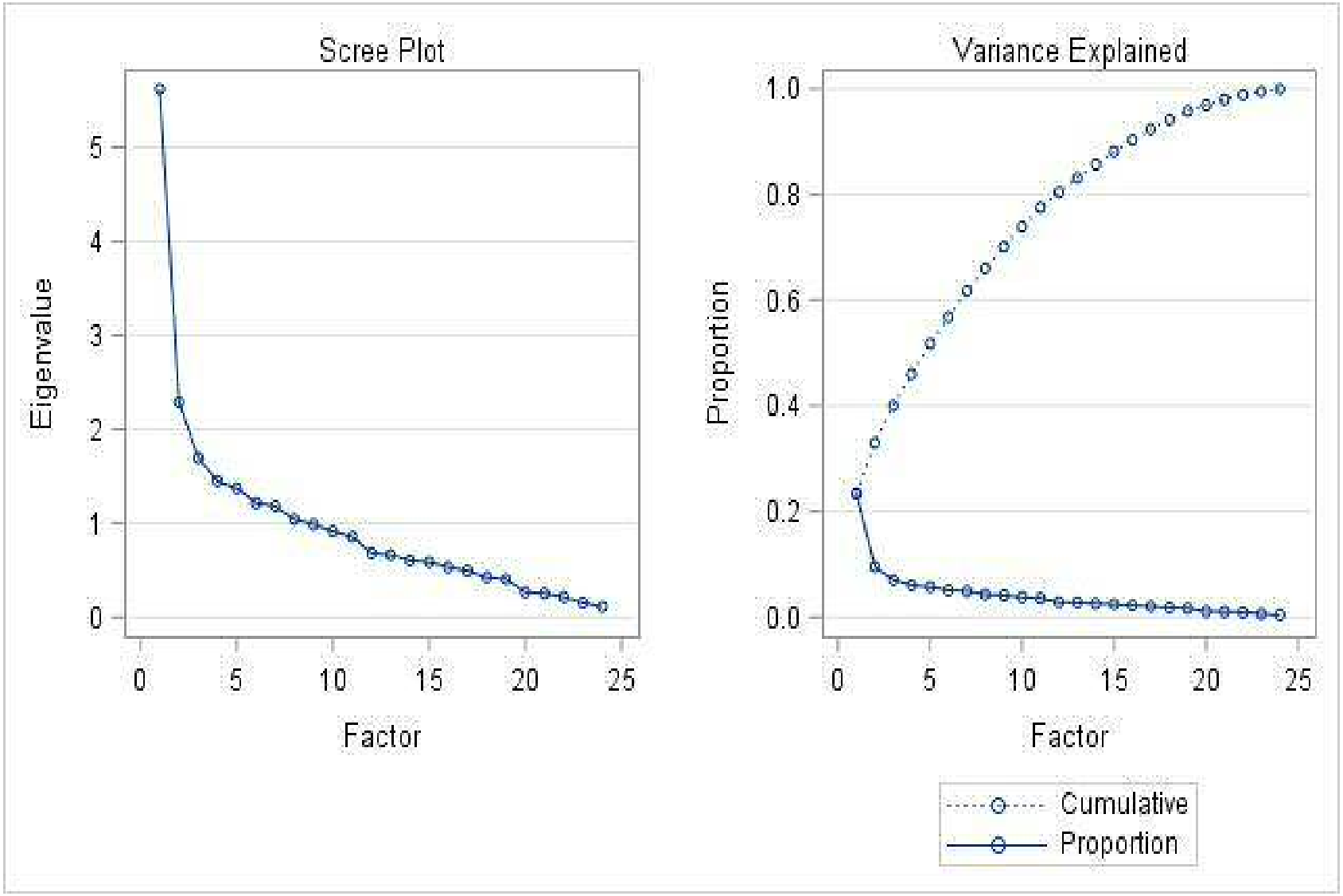
Supplementary Figure 7: Regional plot of TRACK-HD GWAS signal in the MSH3-DHFR region (top, red), along with GTeX eQTL associations with DHFR expression in (top-bottom) whole blood, skeletal muscle, cerebellum, cortex.



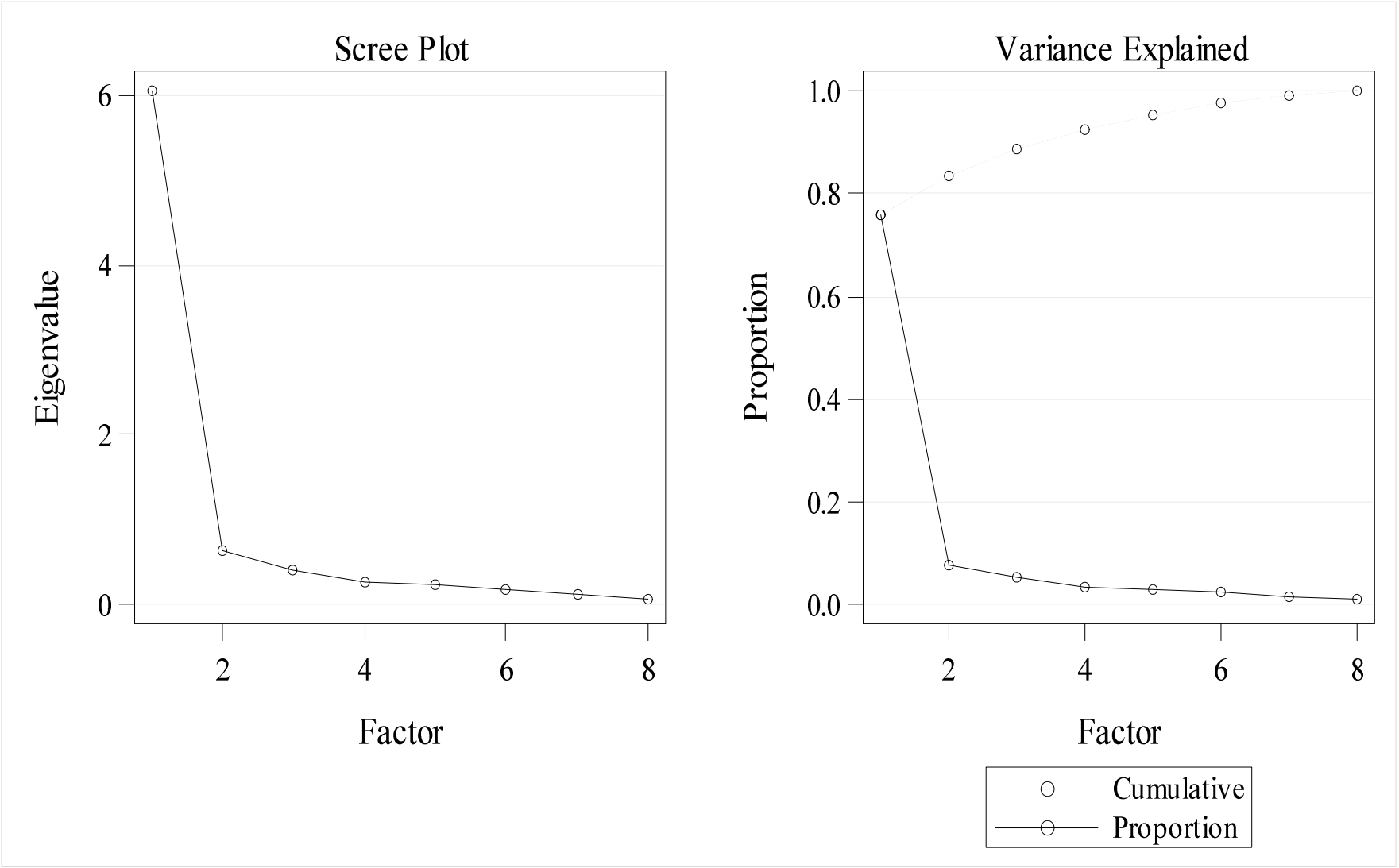
Supplementary Figure 8: Regional plot of REGISTRY GWAS signal in the MSH3-DHFR region (top, blue), along with GTeX eQTL associations with MSH3 expression in (top-bottom) whole blood, transformed fibroblasts.



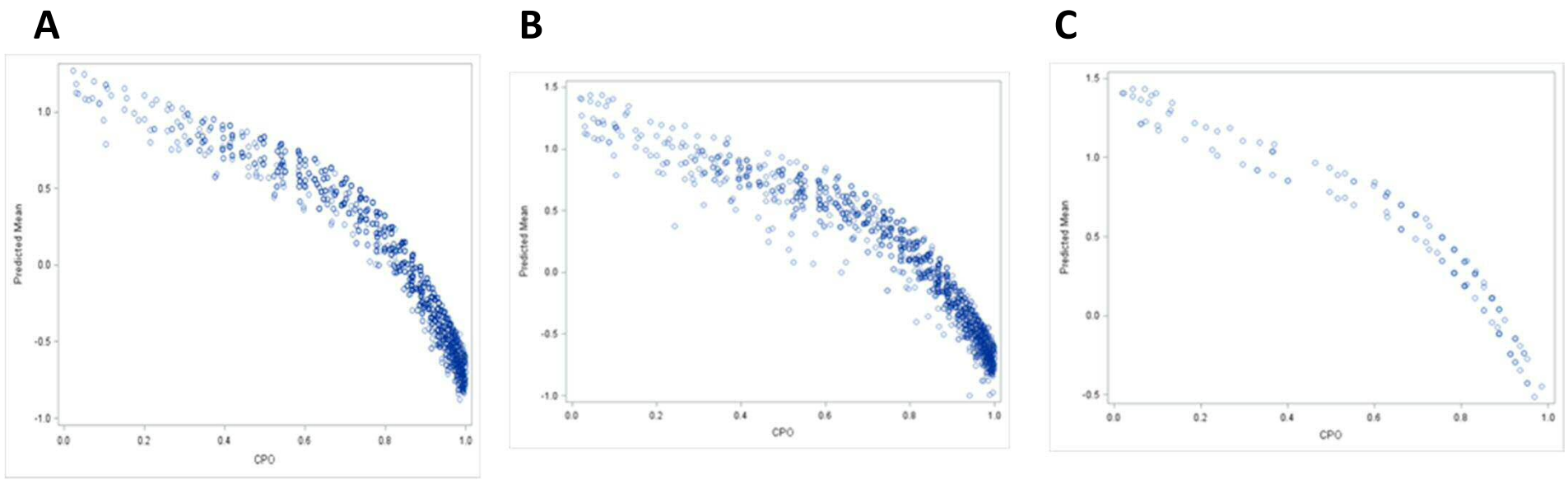
Supplementary Figure 9: (A) Scree Plot and (B) Plot showing proportion of variance explained in the TRACK-HD progression principal component analysis: the dominance of the first PC is illustrated.



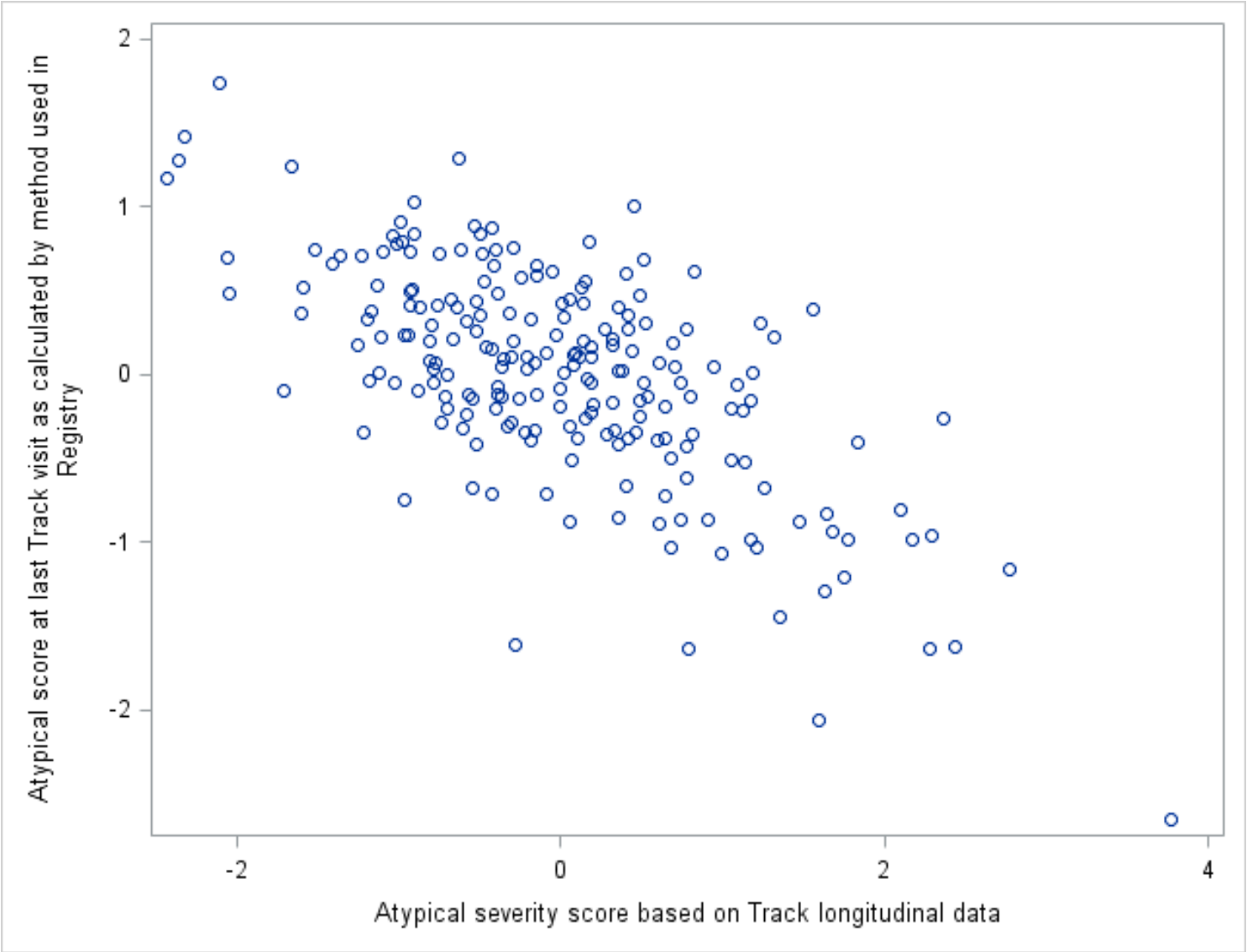
Supplementary Figure 10: (A) Scree Plot and (B) Plot showing proportion of variance explained in the REGISTRY progression principal component analysis: the dominance of the first PC is illustrated.



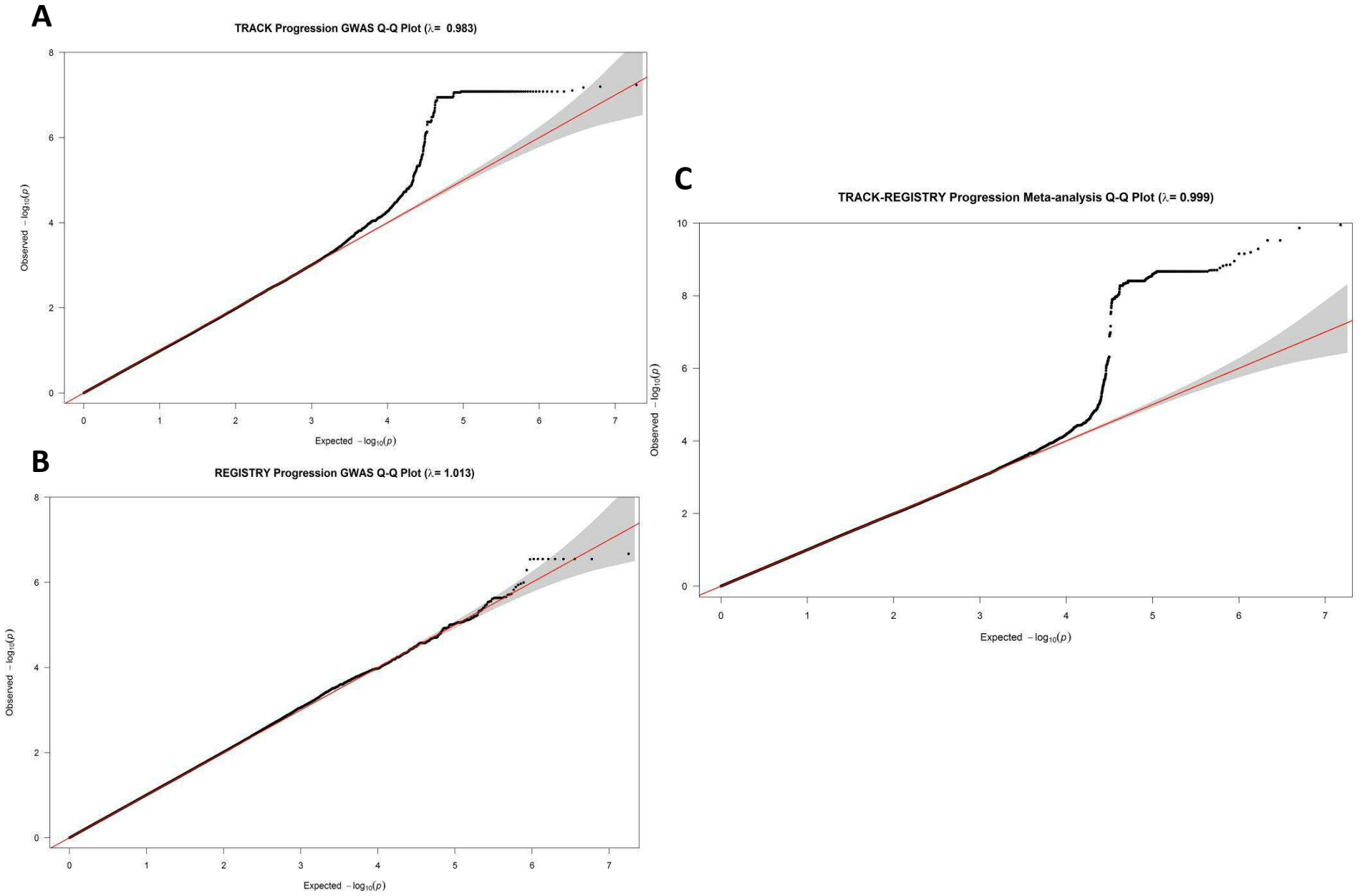
Supplementary Figure 11: Age-CAG severity function against clinical probability of onset (CPO) in REGISTRY. A: plot showing predicted values for all subjects. B: plot of predicted values using only subjects in the CAG 41– 55 range. C: Plot based on extrapolating the severity model to subjects with CAG in the 36-40 range (the appearance of two rather distinct lines are due to the gender effect, with women having lower predicted scores than men).



Supplementary Figure 12: Linear relationship between the longitudinal atypical severity scores used for the TRACK-HD analysis and cross-sectional atypical severity scores at the last TRACK visit when calculated using the method employed for the REGISTRY data (r = .674).



Supplementary Figure 13:
there is no systematic inflation of test statistics.



Supplementary tables:

Supplementary Table 1: Demographic details of TRACK-HD cohort. Further detail can be found in Tabrizi *et al* 2009, 2011, 2012, 2013.

	Number (female)	Age at baseline (years)	CAG repeat length
Manifest	122 (65)	48.0	43.5
Premanifest	96 (53)	40.6	43.0

Supplementary Table 2: List of Variables to be used in TRACK-HD progression analyses. Further detail regarding these measures can be found in Tabrizi *et al* 2009, 2011, 2012, 2013.

Symbol digit modality test (number correct)
Stroop word reading (number correct)
Paced Tapping 3 Hz (inverse std dev)
Spot the Change 5K
Emotion Recognition
Direct Circle (Log annulus length)
Indirect Circle (Log annulus length)
Total brain volume
Ventricular volume
Grey matter volume
White matter volume
Caudate volume
Metronome tapping, nondominant hand
Metronome tapping, nondominant hand
Speeded tapping, nondominant hand
Speeded tapping, nondominant hand
Speeded tapping, nondominant hand
Tongue force—heavy
Tongue force—light
Grip force, dom. hand, heavy condition
Grip force, dom. hand, heavy condition
Grip force, nondom. hand, heavy condition
Grip force, dom. hand, light condition
Grip force, nondom. hand, light condition

Supplementary Table 3: Correlations among Domain-Specific Residual Principal Components in the TRACK-HD analysis, showing that the first principle components of each domain are significantly correlated.

The prefaces "brain", "cog", and "mot" indicate the domain. The suffix f1, f2, etc, numbers the principal components within each domain. Having approximated the residual longitudinal variability within each of the three domains via principal components, we then examined cross-domain relationships among these components. For example, after accounting for CAG-age-risk, testing whether residual longitudinal change in the brain measures correlated with the Q-motor measures.

	brainf1	**brainf2**	**brainf3**	**cogf1**	**cogf2**	**cogf3**	**cogf4**	**motf1**	**motf2**	**motf3**	**motf4**
brainf1	1	0	0	**-0.355**	0.077	0.146	-0.068	**0.43**	0.096	-0.065	-0.139
p	*0*	*1*	*1*	***<.0001***	*0.26*	*0.03*	*0.32*	***<.0001***	*0.16*	*0.34*	*0.04*
brainf2	0	1	0	-0.097	-0.055	0.12	-0.016	0.005	-0.149	-0.043	0.041
p	*1*	*0*	*1*	*0.15*	*0.42*	*0.08*	*0.81*	*0.94*	*0.03*	*0.53*	*0.55*
brainf3	0	0	1	0.016	0.064	0.12	-0.009	0.15	0.05	-0.108	-0.161

p	*1*	*1*	*0*	*0.81*	*0.35*	*0.08*	*0.89*	*0.03*	*0.46*	*0.11*	*0.02*
cogf1				1	0	0	0	**-0.434**	-0.154	0.035	0.112
p				*0*	*1*	*1*	*1*	***<.0001***	*0.02*	*0.6*	*0.09*
cogf2				0	1	0	0	0.035	0.07	-0.12	-0.163
p				*1*	*0*	*1*	*1*	*0.59*	*0.29*	*0.07*	*0.01*
cogf3				0	0	1	0	0.105	-0.017	-0.092	-0.143
p				*1*	*1*	*0*	*1*	*0.11*	*0.8*	*0.16*	*0.03*
cogf4				0	0	0	1	-0.019	-0.05	-0.011	-0.054
p				*1*	*1*	*1*	*0*	*0.77*	*0.44*	*0.87*	*0.42*

Supplementary Table 4: PCA of Residual Longitudinal Change Among Variables form All 3 Domains in the TRACK-HD analysis showing that the variables that correlated with the domain specific analyses also correlated with the common principal component analysis.

Measure	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Symbol Digit	**-0.505**	-0.027	0.135	0.194	0.034	0.047	**-0.394**	-0.121
Stroop Word	**-0.391**	-0.017	0.361	**0.468**	0.078	-0.232	0.087	0.123
Paced Tapping 3 Hz (inverse std dev)	-0.054	-0.123	-0.031	-0.066	0.032	**0.621**	**-0.420**	0.233
Spot the Change 5K	0.224	-0.123	0.113	-0.223	-0.016	0.190	**0.427**	0.479
Emotion Recognition	-0.226	0.188	0.228	0.086	-0.090	**-0.415**	0.098	**0.264**
Direct Circle (Log annulus length)	-0.374	-0.101	**0.419**	0.199	**0.488**	0.258	0.060	-0.027
Indirect Circle (Log annulus length)	**-0.406**	-0.076	**0.407**	0.418	**0.161**	0.336	0.036	0.130
Total brain volume	**0.749**	**-0.457**	0.168	0.077	-0.046	-0.100	-0.115	-0.079
Ventricular volume	**-0.545**	**0.509**	-0.079	-0.125	0.094	0.131	0.274	0.043
Grey matter volume	**0.631**	**-0.491**	0.173	-0.050	-0.088	-0.137	0.038	-0.022
White matter volume	**0.699**	**-0.409**	0.252	-0.085	-0.019	-0.048	0.062	0.044
Caudate volume	**0.584**	**-0.426**	0.082	0.223	0.086	0.083	-0.055	0.046
Metronome tapping, nondominant hand	**0.433**	-0.033	-0.206	-0.338	0.104	0.392	0.037	-0.081
(log of tap initiation SD for all trials)	**0.433**	-0.033	-0.206	-0.338	0.104	0.392	0.037	-0.081
Metronome tapping, nondominant hand (inv tap initiation SD for self-paced trials)	-0.033	-0.212	0.013	0.144	0.116	0.133	0.347	**-0.705**
Speeded tapping, nondominant hand (log of repetition time SD)	0.380	-0.022	**-0.483**	0.315	**0.554**	-0.206	-0.058	0.123
Speeded tapping, nondominant hand (log of tap duration SD)	**0.594**	0.028	-0.335	0.182	**0.437**	-0.061	0.027	0.206
Speeded tapping, nondominant hand (mean intertap time)	0.316	0.373	-0.219	0.006	**0.411**	-0.036	-0.002	-0.120
Tongue force—heavy (log coefficient of variation)	0.147	0.016	-0.332	**0.586**	**-0.445**	0.177	-0.033	0.012
Tongue force—light (log coefficient of variation)	0.247	0.114	-0.399	**0.451**	**-0.407**	0.191	0.217	0.066
Grip force, dom. hand, heavy condition (log of mean orientation)	0.615	**0.488**	0.252	0.009	-0.078	-0.014	-0.336	-0.077
Grip force, dom. hand, heavy condition (log of mean position)	**0.568**	**0.518**	0.207	0.033	-0.027	-0.051	-0.381	-0.042
Grip force, nondom. hand, heavy condition (log of coefficient of variation)	**0.516**	**0.400**	0.213	0.108	0.003	0.122	0.231	-0.145
Grip force, dom. hand, light condition (log of coefficient of variation)	**0.681**	0.311	0.250	0.034	0.016	0.140	0.188	0.114
Grip force, nondom. hand, light condition (log of coefficient of variation)	**0.647**	**0.430**	0.293	0.071	-0.061	0.071	0.163	-0.055
Pct Variance Explained	***23.4***	***9.5***	***7.1***	***6***	***5.7***	***5.1***	***4.9***	***4.3***

Supplementary Table 5: Factor pattern of the first two principal component analysis of the REGISTRY severity score which was used as a progression score for the Registry data. Factor 1 = 1st PC; Factor 2 = 2nd PC.

		Factor Pattern	
Variable	**Variable explanation**	**Factor1**	**Factor2**
sqrtmotor	Square root of the UHDRS total motor score	-0.84233	0.30062
verfl	UHDRS verbal fluency	0.79108	0.24136
sdmt	UHDRS symbol digit score	0.89833	0.1522
scnt	UHDRS Stroop color naming	0.89596	0.25872
swrt	UHDRS Stroop word reading	0.88978	0.2109
sit1	UHDRS Stroop interference score	0.87684	0.21789
tfc	UHDRS total functional capacity	0.8746	-0.39367
fasscore	UHDRS functional assessment scale	0.88355	-0.38555

Supplementary Table 6: Independent association signals from the TRACK-HD Progression GWAS (at p-value < 10-5)

Chromosome	**Start (BP)**	**End (BP)**	**Index SNP (dbSNP b146)**	**Reference Allele (A1)**	**Alternate Allele (A2)**	**Minor Allele Frequency (MAF)**	**INFO score**	**Beta**	**Standard Error**	**P-value**	**Number of SNPs**	**Length (KB)**	**Gene(s) tagged (+/- 20 KB)**
5	79895438	80196258	rs557874766	G	C	0.238	1.000	-0.581	0.107	5.80E-08	380	300.82	DHFR, MSH3, MTRNR2L2
4	74064920	74362359	rs16849472	T	C	0.019	1.000	1.677	0.318	1.34E-07	10	297.44	AFM, AFP, ALB, ANKRD17, LOC728040
3	20860340	20919615	rs111902872	T	C	0.012	0.920	2.419	0.460	1.47E-07	2	59.276	*none*
1	239493679	239917976	rs115206404	A	G	0.009	0.805	2.598	0.503	2.46E-07	2	424.3	CHRM3, CHRM3-AS2
13	89829918	89856005	rs546753686	A	G	0.009	0.949	2.610	0.506	2.50E-07	2	26.088	*none*
6	31892827	31895971	rs188144048	G	C	0.016	1.000	-1.923	0.380	4.30E-07	2	3.145	C2, CFB, LOC102060414
4	52815077	52815077	rs151302971	C	T	0.060	0.998	0.963	0.192	4.98E-07	1	0.001	*none*
10	132818509	132881313	rs150136271	T	C	0.007	0.845	2.881	0.582	7.38E-07	3	62.805	TCERG1L
8	128074135	128092501	rs76712904	T	A	0.009	1.000	2.532	0.512	7.68E-07	13	18.367	PCAT2, PRNCR1
6	147033320	147049507	rs76605780	G	A	0.009	1.000	2.524	0.512	8.42E-07	4	16.188	ADGB
10	24684087	24684087	rs55795540	A	C	0.065	0.995	0.919	0.187	8.85E-07	1	0.001	KIAA1217
22	41330451	41572210	rs185116512	T	C	0.009	0.998	2.482	0.506	9.51E-07	2	241.76	EP300, EP300-AS1, MIR1281, RBX1, XPNPEP3
6	32537468	33038283	rs1062481	T	C	0.009	1.000	2.477	0.512	1.33E-06	6	500.82	BRD2, HLA-DMA, HLA-DMB, HLA-DOA, HLA-DOB, HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DQB2, HLA-DRB1, HLA-DRB6, LOC100294145, PSMB8, PSMB9, TAP1, TAP2, TAPSAR1
11	125957124	126008830	rs200669142	A	AT	0.012	0.999	2.177	0.452	1.44E-06	3	51.707	*none*
2	67244370	67346420	rs56349456	A	G	0.032	0.992	1.334	0.277	1.50E-06	13	102.05	LOC644838, LOC102800447
4	141111891	141111891	rs57282598	G	A	0.009	0.985	2.453	0.510	1.50E-06	1	0.001	*none*

11	4521374	4879985	rs117945252	G	C	0.012	0.840	2.188	0.456	1.60E-06	3	358.61	C11orf40, OR51D1, OR51E1, OR51E2, OR51F1, OR51F2, OR51S1, OR52I1, OR52I2, OR52K1, OR52M1, OR52R1, TRIM68
11	6750653	6917038	rs3889139	A	G	0.007	0.972	2.795	0.584	1.67E-06	4	166.39	GVINP1, OR2AG1, OR2AG2, OR2D2, OR6A2, OR10A2, OR10A4, OR10A5
5	128287222	128679437	rs146180907	T	C	0.009	0.999	2.405	0.505	1.93E-06	6	392.22	ISOC1, MIR4633, SLC27A6
14	78805174	78901796	rs117746737	G	A	0.044	0.860	1.152	0.243	2.10E-06	9	96.623	NRXN3
2	58253090	58774645	rs146045300	G	A	0.009	0.999	1.940	0.412	2.46E-06	5	521.56	FANCL, LINC01122, VRK2
3	36891939	36956117	rs146080846	ATTAT	A	0.014	0.953	1.965	0.418	2.54E-06	2	64.179	TRANK1
19	6675794	6675794	rs183566601	G	C	0.012	0.908	2.139	0.455	2.56E-06	1	0.001	C3, TNFSF14
1	74500787	74665215	rs75274216	G	C	0.016	1.000	1.551	0.331	2.77E-06	2	164.43	FPGT, FPGT-TNNI3K, LRRIQ3
8	58860626	58860626	rs577181066	T	C	0.009	1.000	-2.314	0.495	3.00E-06	1	0.001	*none*
7	153670448	153672048	rs39153	T	A	0.035	1.000	1.264	0.272	3.23E-06	6	1.601	DPP6
9	107972517	107982792	rs33922537	TA	T	0.088	0.991	0.813	0.175	3.25E-06	44	10.276	*none*
2	182396359	182396359	rs185077546	A	G	0.009	1.000	2.344	0.504	3.28E-06	1	0.001	CERKL, ITGA4
2	131955255	132160922	rs377754762	C	G	0.009	0.971	2.342	0.504	3.44E-06	3	205.67	LINC01120, LOC440910, POTEE, WTH3DI
7	37980409	37980409	rs7798464	C	T	0.478	0.999	-0.515	0.111	3.48E-06	1	0.001	EPDR1
6	136578804	136578804	rs182174986	C	T	0.007	1.000	2.694	0.582	3.70E-06	1	0.001	BCLAF1, MTFR2
1	10849645	10849645	rs10864471	G	A	0.076	0.996	0.798	0.173	3.85E-06	1	0.001	CASZ1
16	84533672	85735002	rs544820106	T	C	0.007	0.998	2.692	0.583	3.90E-06	3	1201.3	C16orf74, COTL1, CRISPLD2, FAM92B, GINS2, GSE1, KIAA0513, KLHL36, LINC00311, LOC400548, MIR5093, MIR7851, TLDC1, USP10, ZDHHC7
9	140124101	140326510	rs576074352	T	C	0.009	1.000	2.320	0.505	4.26E-06	3	202.41	C9orf169, C9orf173, ENTPD8, EXD3, FAM166A, LOC100129722, MIR7114, NDOR1, NELFB, NOXA1, NRARP, NSMF, RNF208, RNF224, SLC34A3, TOR4A, TUBB4B
3	147917	147917	rs190854428	T	C	0.014	0.990	1.920	0.418	4.29E-06	1	0.001	*none*
7	20723109	20725291	rs78036476	G	T	0.024	1.000	1.515	0.330	4.44E-06	2	2.183	ABCB5
10	6409502	6409502	rs7915166	C	T	0.308	1.000	0.496	0.108	4.80E-06	1	0.001	*none*
4	178648851	178648851	rs191350537	A	G	0.014	0.999	1.882	0.413	5.15E-06	1	0.001	LINC01098
2	121177685	121177685	rs542948395	T	C	0.009	1.000	2.291	0.504	5.37E-06	1	0.001	*none*
19	40114782	40248603	rs544526021	T	G	0.009	1.000	2.301	0.506	5.42E-06	3	133.82	CLC, LEUTX, LGALS13, LGALS14, LGALS16, LGALS17A, LOC100129935
8	15256069	15266930	rs11203702	A	T	0.118	0.99	0.680	0.15	5.56E-	3	10.8	*none*

							6		0	06		62	
7	126436990	126497288	rs139456699	A	G	0.019	1.000	-1.608	0.354	5.74E-06	3	60.299	GRM8
4	144227742	144227742	rs185067403	A	G	0.009	1.000	2.294	0.506	5.80E-06	1	0.001	*none*
13	20224902	20377448	rs35231784	GC	G	0.260	0.995	-0.495	0.110	6.11E-06	36	152.55	MPHOSPH8, PSPC1
3	71536485	71536485	rs139096029	A	G	0.019	1.000	1.603	0.355	6.26E-06	1	0.001	FOXP1
2	6216990	6265656	rs13017659	A	C	0.068	0.975	0.818	0.181	6.28E-06	4	48.667	*none*
8	141293251	141293251	rs186776689	T	C	0.009	1.000	2.277	0.504	6.30E-06	1	0.001	TRAPPC9
6	117738434	117812254	rs143087465	T	C	0.009	1.000	2.289	0.507	6.42E-06	3	73.821	DCBLD1, ROS1
1	71806741	71806741	rs615589	C	T	0.017	0.953	1.772	0.393	6.45E-06	1	0.001	*none*
1	34835613	34886817	rs10753307	C	G	0.146	1.000	0.646	0.144	6.94E-06	35	51.205	*none*
13	73610584	73624638	rs13378884	G	A	0.280	1.000	0.500	0.111	7.28E-06	2	14.055	KLF5, PIBF1
3	21479214	21521820	rs73045437	A	G	0.131	0.996	0.683	0.152	7.32E-06	2	42.607	ZNF385D
6	24188337	24301530	rs138968896	A	C	0.047	0.998	1.058	0.236	7.39E-06	21	113.19	DCDC2
18	64640322	64640322	rs11663556	T	C	0.009	0.997	2.246	0.505	8.55E-06	1	0.001	*none*
16	21651427	21706726	rs139057628	C	T	0.012	0.881	1.999	0.452	9.82E-06	2	55.3	IGSF6, METTL9, OTOA
11	122679684	122679684	rs5795348	G	GA	0.201	0.987	0.587	0.133	9.91E-06	1	0.001	UBASH3B
10	51520713	51520713	rs74922941	C	T	0.016	1.000	1.701	0.385	9.92E-06	1	0.001	TIMM23B

Supplementary Table 7: Independent association signals from the meta-analysis of TRACK-HD and REGISTRY Progression GWAS (at p-value < 10-5)

Index SNP	P-value	Clump coordinates	Clump size (KB)	Gene(s) tagged
rs1232027	1.12E-10	chr5:79895438..80198404	302.967	DHFR, MSH3, MTRNR2L2
rs73786719	8.53E-07	chr6:147034576..147037984	3.409	ADGB
rs114688092	1.51E-06	chr3:47026101..47315538	289.438	CCDC12, KIF9, KIF9-AS1, KLHL18, NBEAL2, NRADDP, SETD2
rs79029191	1.67E-06	chr18:8053863..8080538	26.676	PTPRM
rs932428	1.79E-06	chr20:37518361..37876772	358.412	DHX35, FAM83D, LOC339568, PPP1R16B
rs3889139	2.13E-06	chr11:6885429..6917038	31.61	OR2D2, OR10A2, OR10A4, OR10A5
rs114643193	2.65E-06	chr4:2844682..2939191	94.51	ADD1, MFSD10, NOP14, NOP14-AS1, SH3BP2
rs6882169	2.72E-06	chr5:167668230..167668230	0.001	CTB-178M22.2, TENM2
rs80260687	2.92E-06	chr8:97232364..97304966	72.603	MTERFD1, PTDSS1, UQCRB
rs28406206	3.13E-06	chr14:105680474..105688082	7.609	BRF1
rs4736525	3.37E-06	chr8:132924474..133030989	106.516	EFR3A, OC90
rs78621558	4.44E-06	chr5:80012735..80012735	0.001	MSH3
rs72715653	4.80E-06	chr4:178641337..178730329	88.993	LINC01098, LINC01099
rs4720024	4.94E-06	chr7:30941255..30942312	1.058	AQP1, FAM188B, INMT-FAM188B
rs117933444	5.75E-06	chr6:167362873..167410443	47.571	FGFR1OP, MIR3939, RNASET2
rs116220136	5.82E-06	chr5:23353255..23436446	83.192	*none*
rs8031584	8.15E-06	chr15:31185616..31292023	106.408	FAN1, MTMR10, TRPM1
rs3013648	9.10E-06	chr13:85296644..85374146	77.503	*none*
rs11197481	9.12E-06	chr10:117708803..117708803	0.001	ATRNL1
rs117440785	9.15E-06	chr10:17411451..17531334	119.884	ST8SIA6, ST8SIA6-AS1
rs111258354	9.87E-06	chr2:60823224..60883232	60.009	*none*

Supplementary Table 8: Co-localisation between TRACK-HD GWAS signal on chromosome 5 and GTeX eQTLs for MSH3, DHFR

Dataset	Dataset source	Most significant eQTL p-value	N Overlapping SNPs	COLOC probability (of shared variants)
MSH3 (Blood)	GTEx	1.70E-28	647	1.76%
MSH3 (Fibroblasts)	GTEx	3.10E-39	646	1.76%

MSH3 (Cerebellum)	GTEx	1.10E-06	592	8.83%
MSH3 (Caudate)	GTEx	1.65E-05	588	25.20%
MSH3 (Cortex)	GTEx	5.53E-05	582	53.10%
DHFR (Blood)	GTEx	5.20E-45	647	98.10%
DHFR (Skeletal muscle)	GTEx	1.30E-68	655	99.20%
DHFR (Cerebellum)	GTEx	7.60E-13	592	28.30%
DHFR (Caudate)	GTEx	2.60E-12	588	99.00%
DHFR (Cortex)	GTEx	4.90E-15	582	96.10%

Supplementary Table 9: Co-localisation between REGISTRY GWAS signal on chromosome 5 and GTeX eQTLs for MSH3, DHFR

Dataset	Dataset source	Most significant eQTL p-value	N Common SNPs	COLOC probability (of shared variants)
MSH3 (Blood)	GTEx	1.70E-28	3289	97.80%
MSH3 (Fibroblasts)	GTEx	3.10E-39	3224	97.80%
MSH3 (Cerebellum)	GTEx	1.10E-06	2888	12.50%
MSH3 (Caudate)	GTEx	1.65E-05	2866	10.40%
MSH3 (Cortex)	GTEx	5.53E-05	2853	23.10%
DHFR (Blood)	GTEx	5.20E-45	3289	36.40%
DHFR (Skeletal muscle)	GTEx	1.30E-68	3336	34.10%
DHFR (Cerebellum)	GTEx	7.60E-13	2888	0.88%
DHFR (Caudate)	GTEx	2.60E-12	2866	43.30%
DHFR (Cortex)	GTEx	4.90E-15	2853	23.10%

Supplementary Table 10: Co-localisation between TRACK-HD GWAS signal on chromosome 5 and GTeX eQTLs for MSH3, DHFR

Dataset	Dataset source	Most significant eQTL p-value	N Overlapping SNPs	COLOC probability (of shared variants)
MSH3 (Blood)	GTEx	1.70E-28	647	1.76%
MSH3 (Fibroblasts)	GTEx	3.10E-39	646	1.76%
MSH3 (Cerebellum)	GTEx	1.10E-06	592	8.83%
MSH3 (Caudate)	GTEx	1.65E-05	588	25.20%
MSH3 (Cortex)	GTEx	5.53E-05	582	53.10%
DHFR (Blood)	GTEx	5.20E-45	647	98.10%
DHFR (Skeletal muscle)	GTEx	1.30E-68	655	99.20%
DHFR (Cerebellum)	GTEx	7.60E-13	592	28.30%
DHFR (Caudate)	GTEx	2.60E-12	588	99.00%
DHFR (Cortex)	GTEx	4.90E-15	582	96.10%

Supplementary Table 11: Independent association signals from the REGISTRY Progression GWAS (at p-value < 10-5)

Chromosome	Start (BP)	End (BP)	Index SNP (dbSNP b146)	Reference Allele (A1)	Alternate Allele (A2)	Minor Allele Frequency (MAF)	INFO score	Beta	Standard Error	P-value	Number of SNPs	Length (KB)	Gene(s) tagged (+/- 20 KB)
10	117708803	117708803	rs11197481	A	G	0.176	0.997	0.193	0.037	2.14E-07	1	0.001	ATRNL1
15	30996093	31314317	rs10611148	A	AAGTT	0.274	0.999	0.160	0.031	2.84E-07	72	318.225	FAN1, HERC2P10, LOC100288637, MTMR10, TRPM1
6	67807895	67905502	rs75695330	C	T	0.268	0.522	0.176	0.034	2.88E-07	12	97.608	*none*

12	117967637	117989548	rs10774933	C	T	0.197	0.992	0.171	0.035	1.08E-06	10	21.912	KSR2
3	86317394	86321260	rs78656706	A	G	0.025	0.610	-0.440	0.091	1.15E-06	2	3.867	*none*
1	151576174	151614297	rs761712980	A	AATAAAT	0.089	0.858	-0.231	0.049	2.21E-06	3	38.124	SNX27
3	93566149	93725515	rs62266135	T	G	0.015	0.500	0.542	0.116	2.77E-06	2	159.367	ARL13B, PROS1, STX19
5	23353255	23436446	rs72754785	G	A	0.045	0.908	0.316	0.067	2.87E-06	4	83.192	*none*
5	36704641	36954077	rs62356368	T	G	0.016	0.985	0.531	0.114	2.92E-06	4	249.437	LOC646719, NIPBL, SLC1A3
20	13209795	13245958	rs759901416	T	TCTCTT	0.156	0.857	0.183	0.039	3.33E-06	3	36.164	ISM1, ISM1-AS1
10	6403262	6407737	rs2387399	T	C	0.358	0.997	0.136	0.029	3.42E-06	2	4.476	*none*
14	33262946	33284981	rs991550	G	A	0.077	0.996	-0.248	0.054	3.60E-06	3	22.036	AKAP6
10	85432343	85432343	rs140550510	G	C	0.014	0.849	0.549	0.119	4.00E-06	1	0.001	*none*
15	92882676	92897269	rs145271683	T	C	0.021	0.908	-0.450	0.098	4.65E-06	4	14.594	*none*
4	3860844	3863228	rs28501173	T	G	0.270	0.997	0.145	0.032	4.66E-06	15	2.385	*none*
12	117075057	117079318	rs144854396	T	TC	0.331	0.890	0.135	0.030	6.01E-06	8	4.262	*none*
16	6945437	6945437	rs188738316	A	G	0.118	0.652	0.205	0.045	6.22E-06	1	0.001	RBFOX1
5	81062170	81062170	rs4703843	G	T	0.165	0.915	0.172	0.038	6.27E-06	1	0.001	SSBP2
11	62532798	62614506	rs41542313	T	C	0.031	0.999	0.367	0.081	6.31E-06	3	81.709	MIR6514, MIR6748, NXF1, POLR2G, SLC3A2, SNHG1, SNORD22, SNORD25, SNORD26, SNORD27, SNORD28, SNORD29, SNORD30, SNORD31, STX5, TAF6L, TMEM179B, TMEM223, WDR74, ZBTB3
21	45715620	45734831	rs3746965	A	G	0.235	1.000	0.150	0.033	6.75E-06	4	19.212	AIRE, C21orf2, PFKL

3	49451639	52028491	rs28587738	A	C	0.014	0.563	0.555	0.124	7.54E-06	5	2576.85	ABHD14A, ABHD14A-ACY1, ABHD14B, ACY1, AMIGO3, AMT, APEH, BSN, BSN-AS2, C3orf18, CACNA2D2, CAMKV, CDHR4, CISH, CYB561D2, DAG1, DOCK3, FAM212A, GMPPB, GNAI2, GNAT1, GPR62, GRM2, HEMK1, HYAL1, HYAL2, HYAL3, IFRD2, IP6K1, IQCF1, IQCF2, IQCF3, IQCF4, IQCF5, IQCF5-AS1, IQCF6, LSMEM2, MANF, MAPKAPK3, MIR4787, MIR5193, MIR5787, MIR6872, MON1A, MST1, MST1R, NAT6, NICN1, NPRL2, PARP3, PCBP4, RAD54L2, RASSF1, RASSF1-AS1, RBM5, RBM5-AS1, RBM6, RBM15B, RHOA, RNF123, RPL29, RRP9, SEMA3B, SEMA3B-AS1, SEMA3F, SLC38A3, TCTA, TEX264, TMEM115, TRAIP, TUSC2, UBA7, VPRBP, ZMYND10
15	31126401	31276476	rs7180337	G	T	0.020	0.621	-0.442	0.099	7.77E-06	22	150.076	FAN1, HERC2P10, MTMR10, TRPM1
15	31345498	31367837	rs28632121	C	T	0.247	0.998	-0.144	0.032	7.96E-06	8	22.34	MIR211, TRPM1
5	158949420	158950938	rs115553365	G	T	0.020	0.799	0.450	0.101	8.58E-06	2	1.519	*none*
19	17164401	17164401	rs73022346	T	G	0.013	0.649	-0.550	0.124	8.93E-06	1	0.001	HAUS8
7	70111666	70238809	rs80237739	C	T	0.025	0.850	-0.405	0.092	9.80E-06	4	127.144	AUTS2

Supplementary Table 12: Gene-wide p-values in TRACK-HD, REGISTRY, the TRACK-REGISTRY meta-analysis and GeM for all genes in the top 14 pathways from GeM

Pathway	Entrez	Gene Symbol	Chr	Start	End	p(TRACK)	p(REGISTRY)	p(META)	p(GeM)	Description
GO:32300	4437	MSH3	5	79950467	80172634	2.94E-08	9.52E-04	8.88E-11	2.03E-02	mismatch repair complex
GO:30983	4437	MSH3	5	79950467	80172634	2.94E-08	9.52E-04	8.88E-11	2.03E-02	mismatched DNA binding
GO:6298	4437	MSH3	5	79950467	80172634	2.94E-08	9.52E-04	8.88E-11	2.03E-02	mismatch repair
KEGG 3430	4437	MSH3	5	79950467	80172634	2.94E-08	9.52E-04	8.88E-11	2.03E-02	KEGG_MISMATCH_REPAIR
KEGG 3430	5425	POLD2	7	44154279	44163169	7.21E-04	3.12E-01	2.75E-03	5.20E-01	KEGG_MISMATCH_REPAIR
KEGG 3430	3978	LIG1	19	48618703	48673560	1.65E-02	8.28E-02	5.35E-04	6.51E-02	KEGG_MISMATCH_REPAIR
KEGG 3430	27030	MLH3	14	75480467	75518235	1.69E-02	6.69E-01	1.47E-01	6.59E-03	KEGG_MISMATCH_REPAIR
GO:6298	27030	MLH3	14	75480467	75518235	1.69E-02	6.69E-01	1.47E-01	6.59E-03	mismatch repair
GO:32407	27030	MLH3	14	75480467	75518235	1.69E-02	6.69E-01	1.47E-01	6.59E-03	MutSalpha complex binding
GO:32300	27030	MLH3	14	75480467	75518235	1.69E-02	6.69E-01	1.47E-01	6.59E-03	mismatch repair complex
GO:30983	27030	MLH3	14	75480467	75518235	1.69E-02	6.69E-01	1.47E-01	6.59E-03	mismatched DNA binding
GO:10822	5534	PPP3R1	2	68405989	68479651	1.82E-02	4.76E-01	6.12E-01	8.40E-01	positive regulation of mitochondrion organization

GO: 33683	2068	ERCC2	19	45854649	45873845	2.03E-02	8.83E-01	3.45E-01	7.45E-01	nucleotide-excision repair, DNA incision
GO: 90200	84334	APOPT1	14	104029299	104057236	2.51E-02	8.19E-01	4.40E-01	8.18E-01	positive regulation of release of cytochrome c from mitochondria
GO: 10822	84334	APOPT1	14	104029299	104057236	2.51E-02	8.19E-01	4.40E-01	8.18E-01	positive regulation of mitochondrion organization
GO: 32389	5395	PMS2	7	6012870	6048737	2.58E-02	3.66E-01	8.84E-03	1.91E-05	MutLalpha complex
GO: 32300	5395	PMS2	7	6012870	6048737	2.58E-02	3.66E-01	8.84E-03	1.91E-05	mismatch repair complex
GO: 30983	5395	PMS2	7	6012870	6048737	2.58E-02	3.66E-01	8.84E-03	1.91E-05	mismatched DNA binding
KEGG 3430	5395	PMS2	7	6012870	6048737	2.58E-02	3.66E-01	8.84E-03	1.91E-05	KEGG_MISMATCH_REPAIR
GO: 6298	5395	PMS2	7	6012870	6048737	2.58E-02	3.66E-01	8.84E-03	1.91E-05	mismatch repair
GO: 32407	5395	PMS2	7	6012870	6048737	2.58E-02	3.66E-01	8.84E-03	1.91E-05	MutSalpha complex binding
GO: 30983	4439	MSH5	6	31707725	31730455	4.35E-02	8.54E-01	7.73E-01	5.14E-01	mismatched DNA binding
GO: 6298	4439	MSH5	6	31707725	31730455	4.35E-02	8.54E-01	7.73E-01	5.14E-01	mismatch repair
KEGG 3430	5982	RFC2	7	73645832	73668738	4.80E-02	5.91E-01	2.02E-02	4.46E-01	KEGG_MISMATCH_REPAIR
GO: 30983	7508	XPC	3	14186647	14220172	5.52E-02	1.04E-01	2.77E-02	5.53E-01	mismatched DNA binding
KEGG 3430	6119	RPA3	7	7676575	7758238	6.55E-02	7.22E-01	9.17E-02	4.40E-01	KEGG_MISMATCH_REPAIR
GO: 32300	4292	MLH1	3	37034841	37092337	6.98E-02	3.97E-04	1.28E-04	4.13E-04	mismatch repair complex
GO: 6298	4292	MLH1	3	37034841	37092337	6.98E-02	3.97E-04	1.28E-04	4.13E-04	mismatch repair
KEGG 3430	4292	MLH1	3	37034841	37092337	6.98E-02	3.97E-04	1.28E-04	4.13E-04	KEGG_MISMATCH_REPAIR
GO: 30983	4292	MLH1	3	37034841	37092337	6.98E-02	3.97E-04	1.28E-04	4.13E-04	mismatched DNA binding
GO: 32407	4292	MLH1	3	37034841	37092337	6.98E-02	3.97E-04	1.28E-04	4.13E-04	MutSalpha complex binding
GO: 32389	4292	MLH1	3	37034841	37092337	6.98E-02	3.97E-04	1.28E-04	4.13E-04	MutLalpha complex
GO: 33683	2067	ERCC1	19	45910591	45927177	7.32E-02	3.96E-01	2.69E-01	3.30E-01	nucleotide-excision repair, DNA incision
GO: 32407	545	ATR	3	142168077	142297668	7.62E-02	7.94E-01	2.71E-01	2.97E-01	MutSalpha complex binding
GO: 90140	79594	MUL1	1	20825941	20834674	8.94E-02	5.22E-01	5.27E-01	4.68E-01	regulation of mitochondrial fission
GO: 90141	79594	MUL1	1	20825941	20834674	8.94E-02	5.22E-01	5.27E-01	4.68E-01	positive regulation of mitochondrial fission
GO: 10822	79594	MUL1	1	20825941	20834674	8.94E-02	5.22E-01	5.27E-01	4.68E-01	positive regulation of mitochondrion organization
GO: 90200	26355	FAM162A	3	122103023	122128961	1.32E-01	7.57E-01	6.93E-01	8.40E-01	positive regulation of release of cytochrome c from mitochondria
GO: 10822	26355	FAM162A	3	122103023	122128961	1.32E-01	7.57E-01	6.93E-01	8.40E-01	positive regulation of mitochondrion organization
GO:1900063	56947	MFF	2	228192228	228222549	1.52E-01	9.63E-01	5.92E-01	3.29E-01	regulation of peroxisome organization
GO: 10822	56947	MFF	2	228192228	228222549	1.52E-01	9.63E-01	5.92E-01	3.29E-01	positive regulation of mitochondrion organization
GO: 90200	56947	MFF	2	228192228	228222549	1.52E-01	9.63E-01	5.92E-01	3.29E-01	positive regulation of release of cytochrome c from mitochondria
GO: 32389	7486	WRN	8	30890778	31031277	1.66E-01	5.59E-01	6.60E-01	3.60E-01	MutLalpha complex
GO: 32300	7486	WRN	8	30890778	31031277	1.66E-01	5.59E-01	6.60E-01	3.60E-01	mismatch repair complex
GO: 10822	637	BID	22	18216906	18257431	1.77E-01	2.99E-02	7.33E-02	2.11E-01	positive regulation of mitochondrion organization
GO: 90200	637	BID	22	18216906	18257431	1.77E-01	2.99E-02	7.33E-02	2.11E-01	positive regulation of release of cytochrome c from mitochondria
GO: 90141	54708	MARCH_5	10	94050920	94113721	1.81E-01	8.26E-03	4.51E-01	5.33E-02	positive regulation of mitochondrial fission
GO: 10822	54708	MARCH_5	10	94050920	94113721	1.81E-01	8.26E-03	4.51E-01	5.33E-02	positive regulation of mitochondrion organization
GO: 90140	54708	MARCH_5	10	94050920	94113721	1.81E-01	8.26E-03	4.51E-01	5.33E-02	regulation of mitochondrial fission
KEGG 3430	29935	RPA4	23	96138907	96140466	1.81E-01	N/A	N/A	N/A	KEGG_MISMATCH_REPAIR
GO: 10822	572	BAD	11	64037300	64052176	1.87E-01	2.48E-01	4.16E-01	1.79E-01	positive regulation of mitochondrion organization
GO: 90200	572	BAD	11	64037300	64052176	1.87E-01	2.48E-01	4.16E-01	1.79E-01	positive regulation of release of cytochrome c from mitochondria
GO: 33683	2071	ERCC3	2	128014866	128051752	1.97E-01	4.27E-01	8.61E-01	7.39E-03	nucleotide-excision repair, DNA incision

GO: 10822	708	C1QBP	17	5336099	5342471	2.05E-01	8.72E-01	2.59E-01	5.99E-01	positive regulation of mitochondrion organization
GO:1900063	57506	MAVS	20	3827446	3856770	2.13E-01	7.14E-02	2.31E-01	8.82E-01	regulation of peroxisome organization
GO: 10822	5366	PMAIP1	18	57567192	57571538	2.38E-01	1.05E-01	2.58E-02	1.10E-01	positive regulation of mitochondrion organization
GO: 90200	5366	PMAIP1	18	57567192	57571538	2.38E-01	1.05E-01	2.58E-02	1.10E-01	positive regulation of release of cytochrome c from mitochondria
GO: 90200	29108	PYCARD	16	31212807	31214097	2.44E-01	4.42E-01	1.57E-01	N/A	positive regulation of release of cytochrome c from mitochondria
GO: 10822	29108	PYCARD	16	31212807	31214097	2.44E-01	4.42E-01	1.57E-01	N/A	positive regulation of mitochondrion organization
GO: 30983	2956	MSH6	2	48010221	48034092	2.46E-01	3.15E-01	1.58E-01	9.36E-02	mismatched DNA binding
GO: 32300	2956	MSH6	2	48010221	48034092	2.46E-01	3.15E-01	1.58E-01	9.36E-02	mismatch repair complex
KEGG 3430	2956	MSH6	2	48010221	48034092	2.46E-01	3.15E-01	1.58E-01	9.36E-02	KEGG_MISMATCH_REPAIR
GO: 6298	2956	MSH6	2	48010221	48034092	2.46E-01	3.15E-01	1.58E-01	9.36E-02	mismatch repair
GO: 10822	51100	SH3GLB1	1	87170253	87213867	2.55E-01	7.63E-01	2.92E-01	5.27E-01	positive regulation of mitochondrion organization
GO: 90141	664	BNIP3	10	133781204	133795435	2.63E-01	1.17E-01	7.70E-01	7.19E-01	positive regulation of mitochondrial fission
GO: 90140	664	BNIP3	10	133781204	133795435	2.63E-01	1.17E-01	7.70E-01	7.19E-01	regulation of mitochondrial fission
GO: 10822	664	BNIP3	10	133781204	133795435	2.63E-01	1.17E-01	7.70E-01	7.19E-01	positive regulation of mitochondrion organization
GO: 90200	664	BNIP3	10	133781204	133795435	2.63E-01	1.17E-01	7.70E-01	7.19E-01	positive regulation of release of cytochrome c from mitochondria
GO: 32407	4595	MUTYH	1	45794914	45806142	2.75E-01	4.31E-01	1.97E-01	1.97E-01	MutSalpha complex binding
GO: 6298	4595	MUTYH	1	45794914	45806142	2.75E-01	4.31E-01	1.97E-01	1.97E-01	mismatch repair
GO: 10822	2810	SFN	1	27189633	27190947	2.78E-01	4.30E-01	2.23E-01	7.65E-01	positive regulation of mitochondrion organization
KEGG 3430	5424	POLD1	19	50887580	50921275	2.84E-01	6.48E-01	6.86E-01	2.11E-01	KEGG_MISMATCH_REPAIR
KEGG 3430	6118	RPA2	1	28218049	28241236	2.94E-01	2.04E-02	1.18E-01	7.45E-01	KEGG_MISMATCH_REPAIR
GO: 6298	2072	ERCC4	16	14014014	14046205	3.00E-01	5.58E-01	2.66E-01	6.21E-01	mismatch repair
GO: 33683	2072	ERCC4	16	14014014	14046205	3.00E-01	5.58E-01	2.66E-01	6.21E-01	nucleotide-excision repair, DNA incision
KEGG 3430	5983	RFC3	13	34392206	34540695	3.15E-01	7.80E-01	7.18E-01	6.12E-01	KEGG_MISMATCH_REPAIR
GO: 90200	7157	TP53	17	7571720	7590868	3.21E-01	5.79E-01	2.20E-01	2.47E-01	positive regulation of release of cytochrome c from mitochondria
GO: 10822	7157	TP53	17	7571720	7590868	3.21E-01	5.79E-01	2.20E-01	2.47E-01	positive regulation of mitochondrion organization
GO: 10822	207	AKT1	14	105235686	105262080	3.62E-01	4.10E-01	5.64E-01	3.96E-01	positive regulation of mitochondrion organization
KEGG 3430	5981	RFC1	4	39289069	39368001	3.64E-01	6.29E-01	7.60E-01	6.19E-01	KEGG_MISMATCH_REPAIR
GO: 90200	581	BAX	19	49458117	49465055	3.65E-01	1.25E-01	2.47E-01	8.13E-01	positive regulation of release of cytochrome c from mitochondria
GO: 10822	581	BAX	19	49458117	49465055	3.65E-01	1.25E-01	2.47E-01	8.13E-01	positive regulation of mitochondrion organization
GO: 90200	90427	BMF	15	40380091	40401075	3.71E-01	5.25E-02	3.21E-02	5.08E-01	positive regulation of release of cytochrome c from mitochondria
GO: 10822	90427	BMF	15	40380091	40401075	3.71E-01	5.25E-02	3.21E-02	5.08E-01	positive regulation of mitochondrion organization
GO: 10822	10891	PPARGC1A	4	23793644	23891700	3.79E-01	1.49E-01	1.47E-01	3.43E-01	positive regulation of mitochondrion organization
GO: 10822	65018	PINK1	1	20959948	20978004	3.83E-01	8.71E-01	5.33E-01	4.83E-01	positive regulation of mitochondrion organization
GO: 90200	65018	PINK1	1	20959948	20978004	3.83E-01	8.71E-01	5.33E-01	4.83E-01	positive regulation of release of cytochrome c from mitochondria
GO: 90200	10962	MLLT11	1	151032151	151040973	3.90E-01	7.62E-01	9.23E-01	4.75E-01	positive regulation of release of cytochrome c from mitochondria
GO: 10822	10962	MLLT11	1	151032151	151040973	3.90E-01	7.62E-01	9.23E-01	4.75E-01	positive regulation of mitochondrion organization
GO: 32300	4436	MSH2	2	47630206	47710367	3.98E-01	3.10E-01	7.03E-01	5.49E-01	mismatch repair complex
GO: 30983	4436	MSH2	2	47630206	47710367	3.98E-01	3.10E-01	7.03E-01	5.49E-01	mismatched DNA binding
GO: 6298	4436	MSH2	2	47630206	47710367	3.98E-01	3.10E-01	7.03E-01	5.49E-01	mismatch repair

KEGG 3430	4436	MSH2	2	47630206	47710367	3.98E-01	3.10E-01	7.03E-01	5.49E-01	KEGG_MISMATCH_REPAIR
GO: 10822	841	CASP8	2	202098166	202152434	4.15E-01	8.81E-01	4.49E-01	3.35E-01	positive regulation of mitochondrion organization
GO: 10822	7533	YWHAH	22	32340479	32353590	4.25E-01	7.16E-01	2.86E-01	6.25E-01	positive regulation of mitochondrion organization
GO: 10822	8655	DYNLL1	12	120907660	120936298	4.50E-01	3.11E-01	4.07E-01	4.21E-01	positive regulation of mitochondrion organization
GO: 32407	5378	PMS1	2	190648811	190742355	4.57E-01	8.23E-01	3.36E-01	7.24E-02	MutSalpha complex binding
GO: 32389	5378	PMS1	2	190648811	190742355	4.57E-01	8.23E-01	3.36E-01	7.24E-02	MutLalpha complex
GO: 32300	5378	PMS1	2	190648811	190742355	4.57E-01	8.23E-01	3.36E-01	7.24E-02	mismatch repair complex
GO: 30983	5378	PMS1	2	190648811	190742355	4.57E-01	8.23E-01	3.36E-01	7.24E-02	mismatched DNA binding
GO: 6298	5378	PMS1	2	190648811	190742355	4.57E-01	8.23E-01	3.36E-01	7.24E-02	mismatch repair
GO: 33683	2073	ERCC5	13	103498191	103528351	4.73E-01	7.10E-01	3.43E-01	2.62E-01	nucleotide-excision repair, DNA incision
GO: 10822	7755	ZNF205	16	3162563	3170518	4.74E-01	9.01E-01	7.24E-01	9.47E-01	positive regulation of mitochondrion organization
GO:90200	8743	TNFSF10	3	172223298	172241297	4.77E-01	6.95E-01	6.77E-01	6.09E-01	positive regulation of release of cytochrome c from mitochondria
GO:10822	8743	TNFSF10	3	172223298	172241297	4.77E-01	6.95E-01	6.77E-01	6.09E-01	positive regulation of mitochondrion organization
KEGG 3430	6742	SSBP1	7	141438121	141450288	4.81E-01	8.18E-01	8.67E-01	5.17E-01	KEGG_MISMATCH_REPAIR
GO:10822	28958	COA3	17	40949652	40950704	4.87E-01	1.75E-03	5.11E-01	N/A	positive regulation of mitochondrion organization
GO:6298	11277	TREX1	3	48506919	48509044	4.91E-01	4.76E-01	7.94E-01	4.11E-01	mismatch repair
GO:32407	11277	TREX1	3	48506919	48509044	4.91E-01	4.76E-01	7.94E-01	4.11E-01	MutSalpha complex binding
GO:33683	22909	FAN1	15	31196055	31235311	5.30E-01	2.16E-06	1.15E-04	2.10E-09	nucleotide-excision repair, DNA incision
GO:10822	10572	SIVA1	14	105219470	105225996	5.32E-01	1.48E-01	6.74E-01	8.89E-01	positive regulation of mitochondrion organization
GO:6298	9156	EXO1	1	242011493	242053241	5.56E-01	9.35E-01	9.03E-01	2.23E-01	mismatch repair
KEGG 3430	9156	EXO1	1	242011493	242053241	5.56E-01	9.35E-01	9.03E-01	2.23E-01	KEGG_MISMATCH_REPAIR
GO: 90200	10105	PPIF	10	81107220	81115090	5.62E-01	2.02E-01	4.28E-01	4.88E-01	positive regulation of release of cytochrome c from mitochondria
GO: 10822	10105	PPIF	10	81107220	81115090	5.62E-01	2.02E-01	4.28E-01	4.88E-01	positive regulation of mitochondrion organization
GO: 6298	7161	TP73	1	3569129	3652765	5.69E-01	3.18E-01	4.40E-01	5.54E-01	mismatch repair
GO:10822	7531	YWHAE	17	1247834	1303556	5.70E-01	8.16E-01	4.96E-01	5.15E-01	positive regulation of mitochondrion organization
GO: 10822	7532	YWHAG	7	75956108	75988342	5.78E-01	4.82E-01	9.74E-01	8.36E-02	positive regulation of mitochondrion organization
GO: 10822	7534	YWHAZ	8	101930804	101965623	5.89E-01	1.51E-01	1.89E-01	5.93E-02	positive regulation of mitochondrion organization
GO: 90140	64423	INF2	14	105155943	105185947	5.93E-01	2.11E-01	2.83E-01	5.52E-01	regulation of mitochondrial fission
GO: 10822	578	BAK1	6	33540323	33548070	5.98E-01	7.98E-01	7.78E-01	3.03E-01	positive regulation of mitochondrion organization
GO: 90200	578	BAK1	6	33540323	33548070	5.98E-01	7.98E-01	7.78E-01	3.03E-01	positive regulation of release of cytochrome c from mitochondria
GO: 33683	4913	NTHL1	16	2089816	2097867	6.25E-01	5.50E-01	4.66E-01	6.35E-01	nucleotide-excision repair, DNA incision
GO: 90200	10018	BCL2L11	2	111878491	111926022	6.27E-01	8.58E-01	8.05E-01	1.51E-02	positive regulation of release of cytochrome c from mitochondria
GO: 10822	10018	BCL2L11	2	111878491	111926022	6.27E-01	8.58E-01	8.05E-01	1.51E-02	positive regulation of mitochondrion organization
GO: 10822	4836	NMT1	17	43138680	43186384	6.37E-01	9.42E-01	9.35E-01	4.65E-01	positive regulation of mitochondrion organization
GO: 10822	10971	YWHAQ	2	9724106	9771106	6.38E-01	1.92E-01	6.28E-01	7.69E-01	positive regulation of mitochondrion organization
GO: 10822	7529	YWHAB	20	43514344	43537161	6.50E-01	2.53E-01	4.98E-01	8.31E-01	positive regulation of mitochondrion organization
GO: 6298	10714	POLD3	11	74303575	74354105	6.51E-01	8.79E-01	6.36E-01	1.52E-01	mismatch repair
KEGG 3430	10714	POLD3	11	74303575	74354105	6.51E-01	8.79E-01	6.36E-01	1.52E-01	KEGG_MISMATCH_REPAIR
GO: 30983	6996	TDG	12	104359593	104382656	6.84E-01	1.83E-01	2.10E-01	4.78E-01	mismatched DNA binding

GO: 6298	6996	TDG	12	104359593	104382656	6.84E-01	1.83E-01	2.10E-01	4.78E-01	mismatch repair
GO: 90140	1723	DHODH	16	72042643	72059316	6.96E-01	9.59E-01	7.30E-01	4.85E-01	regulation of mitochondrial fission
GO: 6298	25	ABL1	9	133589268	133763062	6.97E-01	6.47E-01	9.21E-01	1.81E-01	mismatch repair
GO: 30983	4438	MSH4	1	76262556	76378923	7.24E-01	2.05E-01	2.13E-01	1.40E-01	mismatched DNA binding
KEGG 3430	5985	RFC5	12	118454506	118470044	7.38E-01	1.15E-01	2.33E-01	3.95E-01	KEGG_MISMATCH_REPAIR
GO:190006 3	1005 9	DNM1L	12	32832137	32898584	7.55E-01	8.32E-01	6.94E-01	1.36E-03	regulation of peroxisome organization
GO: 90141	1005 9	DNM1L	12	32832137	32898584	7.55E-01	8.32E-01	6.94E-01	1.36E-03	positive regulation of mitochondrial fission
GO: 90200	1005 9	DNM1L	12	32832137	32898584	7.55E-01	8.32E-01	6.94E-01	1.36E-03	positive regulation of release of cytochrome c from mitochondria
GO: 10822	1005 9	DNM1L	12	32832137	32898584	7.55E-01	8.32E-01	6.94E-01	1.36E-03	positive regulation of mitochondrion organization
GO: 90140	1005 9	DNM1L	12	32832137	32898584	7.55E-01	8.32E-01	6.94E-01	1.36E-03	regulation of mitochondrial fission
KEGG 3430	6117	RPA1	17	1733273	1802848	7.75E-01	2.96E-01	5.51E-01	4.76E-01	KEGG_MISMATCH_REPAIR
GO: 10822	5533	PPP3CC	8	22298483	22398657	7.99E-01	4.58E-01	7.29E-01	3.38E-01	positive regulation of mitochondrion organization
KEGG 3430	5984	RFC4	3	186507681	186524484	8.08E-01	7.04E-01	7.95E-01	3.01E-01	KEGG_MISMATCH_REPAIR
GO: 4748	6240	RRM1	11	4115924	4160106	8.20E-01	6.60E-01	9.85E-01	3.40E-01	ribonucleoside-diphosphate reductase activity, thioredoxin disulfide as acceptor
GO: 16728	6240	RRM1	11	4115924	4160106	8.20E-01	6.60E-01	9.85E-01	3.40E-01	oxidoreductase activity, acting on CH or CH2 groups, disulfide as acceptor
GO: 30983	5111	PCNA	20	5095599	5107268	8.29E-01	2.76E-01	6.40E-01	3.55E-01	mismatched DNA binding
KEGG 3430	5111	PCNA	20	5095599	5107268	8.29E-01	2.76E-01	6.40E-01	3.55E-01	KEGG_MISMATCH_REPAIR
GO: 6298	5111	PCNA	20	5095599	5107268	8.29E-01	2.76E-01	6.40E-01	3.55E-01	mismatch repair
GO: 90200	638	BIK	22	43506754	43525718	8.52E-01	6.42E-01	8.52E-01	1.19E-01	positive regulation of release of cytochrome c from mitochondria
GO: 10822	638	BIK	22	43506754	43525718	8.52E-01	6.42E-01	8.52E-01	1.19E-01	positive regulation of mitochondrion organization
GO: 10822	596	BCL2	18	60790579	60986613	8.65E-01	5.93E-01	4.81E-01	6.54E-01	positive regulation of mitochondrion organization
GO: 10822	3002	GZMB	14	25100160	25103432	8.84E-01	8.26E-01	8.18E-01	6.33E-01	positive regulation of mitochondrion organization
GO: 10822	2711 3	BBC3	19	47724079	47736023	8.89E-01	4.98E-01	7.87E-01	2.78E-01	positive regulation of mitochondrion organization
GO: 90200	2711 3	BBC3	19	47724079	47736023	8.89E-01	4.98E-01	7.87E-01	2.78E-01	positive regulation of release of cytochrome c from mitochondria
GO: 16728	6241	RRM2	2	10262695	10271546	8.96E-01	3.35E-01	3.69E-01	2.65E-01	oxidoreductase activity, acting on CH or CH2 groups, disulfide as acceptor
GO: 4748	6241	RRM2	2	10262695	10271546	8.96E-01	3.35E-01	3.69E-01	2.65E-01	ribonucleoside-diphosphate reductase activity, thioredoxin disulfide as acceptor
GO: 10822	8398	PLA2G6	22	38507502	38577836	9.01E-01	2.91E-01	6.64E-01	1.80E-01	positive regulation of mitochondrion organization
GO: 90200	8398	PLA2G6	22	38507502	38577836	9.01E-01	2.91E-01	6.64E-01	1.80E-01	positive regulation of release of cytochrome c from mitochondria
GO: 90200	8739	HRK	12	117299027	117319232	9.10E-01	6.48E-01	8.21E-01	4.30E-01	positive regulation of release of cytochrome c from mitochondria
GO: 10822	8739	HRK	12	117299027	117319232	9.10E-01	6.48E-01	8.21E-01	4.30E-01	positive regulation of mitochondrion organization
GO: 10822	5599	MAPK8	10	49609687	49643183	9.32E-01	7.42E-01	8.49E-01	7.87E-01	positive regulation of mitochondrion organization
GO: 4748	5048 4	RRM2B	8	103216729	103251346	9.38E-01	6.29E-01	8.45E-01	6.44E-06	ribonucleoside-diphosphate reductase activity, thioredoxin disulfide as acceptor
GO: 16728	5048 4	RRM2B	8	103216729	103251346	9.38E-01	6.29E-01	8.45E-01	6.44E-06	oxidoreductase activity, acting on CH or CH2 groups, disulfide as acceptor
GO: 10822	1407 35	DYNLL2	17	56160780	56167618	9.58E-01	8.08E-01	8.19E-01	8.93E-01	positive regulation of mitochondrion organization
KEGG 3430	5780 4	POLD4	11	67118236	67121067	9.59E-01	6.48E-01	9.21E-01	3.74E-01	KEGG_MISMATCH_REPAIR
GO: 10822	8470 9	MGARP	4	140187317	140201492	9.78E-01	8.81E-01	8.98E-01	1.51E-01	positive regulation of mitochondrion organization

Supplementary Table 13: Setscreen enrichment p-values for the Pearl et al. (2015) pathways in TRACK-HD, REGISTRY, the TRACK-HD meta-analysis and GeM

Gene Set	p(TRACK)	p(REGISTRY)	p(META)	p (GeM)	Description1	Description2	Description3	Description4
2071015	9.05E-07	4.43E-03	2.93E-11	2.01E-02	Repair_pathway	SSR	MMR	Mismatch_and_loop_recognition_factors
2071000	2.43E-06	6.85E-02	1.49E-14	5.15E-04	Repair_pathway	SSR	MMR	
2070000	5.77E-03	4.76E-02	3.32E-07	1.42E-02	Repair_pathway	SSR		
2071017	1.95E-02	2.44E-02	5.84E-05	8.92E-08	Repair_pathway	SSR	MMR	MutL_homologs
2111513	4.71E-02	2.55E-01	8.12E-01	2.86E-03	Repair_pathway	Associated_process	TLS	DNA_polymerases
2070600	5.02E-02	7.99E-01	1.10E-01	2.92E-01	Repair_pathway	SSR	NER	
2070607	5.18E-02	7.61E-01	3.02E-02	2.26E-01	Repair_pathway	SSR	NER	TCR_(Transcription_coupled_repair)
2071104	5.35E-02	3.90E-01	2.07E-02	5.37E-02	Repair_pathway	SSR	BER	LONG_PATCH-BER_factors
2022100	6.69E-02	3.19E-02	7.21E-04	7.29E-02	Repair_pathway	DSR	Alt-NHEJ	
1100000	7.52E-02	6.14E-01	1.94E-01	6.13E-01	Associated_process	DNA_replication		
1080700	8.99E-02	8.35E-01	2.82E-01	4.92E-01	Associated_process	Checkpoint_factors	S-CC_phase	
1051930	1.02E-01	5.68E-01	1.30E-01	7.62E-01	Associated_process	Ubiquitin_response	Ubiquitin-_conjugating_enzymes_(E2)	UBL-conjugating_enzymes
2000000	1.13E-01	2.60E-01	1.03E-03	1.11E-02	Repair_pathway			
2070605	1.14E-01	5.00E-01	8.14E-01	4.64E-01	Repair_pathway	SSR	NER	DNA_polymerase_epsilon
1030000	1.59E-01	1.90E-01	3.59E-01	2.63E-01	Associated_process	Telomere_maintenance		
2070606	1.60E-01	9.56E-01	6.55E-01	5.49E-01	Repair_pathway	SSR	NER	DNA_polymerase_kappa
2071020	1.73E-01	3.14E-01	9.86E-03	7.97E-02	Repair_pathway	SSR	MMR	Other_MMR_factors
1051900	1.97E-01	7.69E-01	1.71E-01	8.19E-01	Associated_process	Ubiquitin_response	Ubiquitin-_conjugating_enzymes_(E2)	
2071023	2.15E-01	1.73E-01	7.67E-02	5.90E-01	Repair_pathway	SSR	MMR	RPA_(replication_factor_A)
1081300	2.15E-01	8.71E-01	4.25E-01	6.96E-01	Associated_process	Checkpoint_factors	HRAD17(Rad24)-_RFC_complex	
1051208	2.41E-01	2.50E-01	3.12E-01	5.81E-01	Associated_process	Ubiquitin_response	Ubiquitin_ligases_(E3)	single_Ring-finger_type_E3
1080900	2.50E-01	4.77E-01	9.41E-01	2.74E-01	Associated_process	Checkpoint_factors	G1-S_checkpoint	
2071003	2.58E-01	8.68E-01	3.40E-01	1.57E-01	Repair_pathway	SSR	MMR	DNA_polymerase_delta
1051222	2.87E-01	2.82E-01	1.50E-01	6.61E-01	Associated_process	Ubiquitin_response	Ubiquitin_ligases_(E3)	Riddle_syndrome!
1080800	2.87E-01	3.88E-01	7.69E-01	2.52E-01	Associated_process	Checkpoint_factors	G1-CC_phase	
2070603	2.92E-01	8.34E-01	5.37E-01	4.50E-01	Repair_pathway	SSR	NER	DNA_polymerase_delta
2071010	2.92E-01	7.60E-01	6.37E-01	7.12E-01	Repair_pathway	SSR	MMR	RFC_(replication_factor_C)
1051221	3.18E-01	1.56E-01	1.06E-02	2.79E-01	Associated_process	Ubiquitin_response	Ubiquitin_ligases_(E3)	Other_single_Ring-_finger_type_E3
1010000	3.23E-01	4.39E-01	3.23E-01	8.30E-01	Associated_process	Chromatin_remodelling		
1051829	3.28E-01	5.91E-01	5.58E-01	9.17E-01	Associated_process	Ubiquitin_response	Ubiquitin-_activating_enzymes_(E1)	UBL-activating_enzymes
1051800	3.29E-01	5.91E-01	5.58E-01	9.17E-01	Associated_process	Ubiquitin_response	Ubiquitin-_activating_enzymes_(E1)	
1051927	3.31E-01	7.89E-01	4.15E-01	6.74E-01	Associated_process	Ubiquitin_response	Ubiquitin-_conjugating_enzymes_(E2)	Ubiquitin-conjugating_enzymes
3060000	3.41E-01	1.70E-01	3.61E-01	7.39E-01	Genes_with_probable_DDR_role	Direct_Repair_(not_in_humans)		

1031600	3.86E-01	8.44E-01	5.12E-01	6.69E-01	Associated_process	Telomere_maintenance	Alternative_mechanism	
1031616	3.86E-01	8.44E-01	5.12E-01	6.69E-01	Associated_process	Telomere_maintenance	Alternative_mechanism	MRN_Complex
2020200	4.09E-01	6.98E-01	5.00E-01	4.77E-01	Repair_pathway	DSR	HR_(Homologous_Recombination)	
1052000	4.20E-01	3.11E-01	4.35E-01	8.24E-01	Associated_process	Ubiquitin_response	Ubiquitins_and_Ubiquitin-like_proteins	
1052028	4.20E-01	3.11E-01	4.35E-01	8.24E-01	Associated_process	Ubiquitin_response	Ubiquitins_and_Ubiquitin-like_proteins	Ubiquitins
1000000	4.26E-01	4.38E-01	5.76E-01	3.21E-01	Associated_process			
1082500	4.29E-01	1.79E-01	5.65E-01	6.91E-01	Associated_process	Checkpoint_factors	FPC_(fork_protection_complex)	
2111531	4.30E-01	2.91E-01	2.99E-01	7.84E-01	Repair_pathway	Associated_process	TLS	Y-family_DNA_polymerases
2071018	4.44E-01	2.64E-01	2.34E-01	1.12E-01	Repair_pathway	SSR	MMR	MutS_homologs_specialized_for_meiosis
2110000	4.48E-01	4.49E-01	5.96E-01	4.80E-02	Repair_pathway	Associated_process		
2111500	4.48E-01	4.49E-01	5.96E-01	4.80E-02	Repair_pathway	Associated_process	TLS	
2020000	4.71E-01	4.39E-01	8.35E-02	4.20E-02	Repair_pathway	DSR		
1050500	4.76E-01	8.55E-01	8.56E-01	7.18E-01	Associated_process	Ubiquitin_response	Deubiquitinating_enzyme_(DUB)	
1050501	4.76E-01	8.55E-01	8.56E-01	7.18E-01	Associated_process	Ubiquitin_response	Deubiquitinating_enzyme_(DUB)	UBL-specific_proteases_(ULPs)
1080000	4.86E-01	4.50E-01	8.20E-01	2.85E-01	Associated_process	Checkpoint_factors		
2072800	4.97E-01	5.82E-01	7.02E-02	3.98E-02	Repair_pathway	SSR	Other_SSR_genes	
2020400	5.07E-01	7.84E-01	8.18E-01	5.80E-01	Repair_pathway	DSR	NHEJ	
2071100	5.18E-01	1.14E-01	2.76E-01	1.65E-01	Repair_pathway	SSR	BER	
1082600	5.20E-01	5.64E-01	6.17E-01	5.95E-01	Associated_process	Checkpoint_factors	G2-CC_phase	
1090000	5.70E-01	5.67E-01	6.15E-01	6.62E-01	Associated_process	p53_pathway		
1050000	5.88E-01	3.44E-01	2.17E-01	7.47E-01	Associated_process	Ubiquitin_response		
2070602	5.93E-01	1.61E-01	3.08E-01	5.35E-01	Repair_pathway	SSR	NER	GGR_(Global_genome_repair)
2020300	6.05E-01	5.24E-01	6.24E-01	8.22E-01	Repair_pathway	DSR	Other_DSR_genes	
2071119	6.09E-01	6.72E-02	9.07E-01	2.64E-01	Repair_pathway	SSR	BER	Other_BER_factors
2071111	6.11E-01	2.27E-01	5.24E-01	9.70E-01	Repair_pathway	SSR	BER	AP_endonucleases
1082700	6.14E-01	6.85E-01	9.25E-01	1.51E-01	Associated_process	Checkpoint_factors	G2-M_checkpoint	
2021400	6.22E-01	4.96E-02	1.45E-01	9.20E-01	Repair_pathway	DSR	HR_(Homologous Recombination)	
1051700	6.42E-01	4.61E-01	5.63E-01	1.52E-01	Associated_process	Ubiquitin_response	Ubiquitin-like_proteins_(UBLs)	
1051725	6.42E-01	4.61E-01	5.63E-01	1.52E-01	Associated_process	Ubiquitin_response	Ubiquitin-like_proteins_(UBLs)	SUMO
1051200	6.61E-01	7.44E-02	5.58E-02	3.70E-01	Associated_process	Ubiquitin_response	Ubiquitin_ligases_(E3)	
1082900	6.63E-01	8.72E-01	8.87E-01	4.58E-01	Associated_process	Checkpoint_factors	Rad17-Mec3-_Ddc1_complex	
1082200	6.69E-01	8.04E-02	2.30E-01	2.61E-01	Associated_process	Checkpoint_factors	damage_in_S_phase	
2111514	7.20E-01	5.25E-01	7.10E-01	4.37E-01	Repair_pathway	Associated_process	TLS	epistasis_group

2020100	7.23E-01	5.41E-01	5.70E-01	2.93E-04	Repair_pathway	DSR	FA_(Fanconi_anemia_pathway)	
1040000	7.46E-01	5.93E-01	6.62E-01	3.78E-01	Associated_process	Chromosome_segregation		
3000000	7.86E-01	6.19E-01	3.00E-01	7.39E-01	Genes_with_probable_DDR_role			
2072300	7.97E-01	3.24E-01	8.88E-01	8.75E-01	Repair_pathway	SSR	Direct_Repair	
2072400	8.27E-01	3.89E-03	6.87E-02	2.76E-01	Repair_pathway	SSR	DNA_replication	
2071124	8.39E-01	8.94E-01	8.19E-01	3.16E-01	Repair_pathway	SSR	BER	SHORT_PATCH-BER_factors
2071112	9.02E-01	1.67E-01	3.58E-01	5.51E-01	Repair_pathway	SSR	BER	DNA_glycosylases
1051209	9.25E-01	1.38E-02	5.59E-02	7.56E-01	Associated_process	Ubiquitin_response	Ubiquitin_ligases_(E3)	single_Ring-finger_type_E4
1120000	9.58E-01	6.23E-01	9.97E-01	6.78E-05	Associated_process	Modulation_of_nucleotide_pools		
1083000	9.62E-01	7.83E-01	9.16E-01	8.57E-01	Associated_process	Checkpoint_factors	RAD9-Hus1-Rad1_complex	

Supplementary Table 14: Gene-wide p-values for the most significant genes in the two Pearl et al. pathways showing significant enrichment in TRACK

Entrez	Gene Symbol	Chr	Start	End	p(TRACK)	p(REG)	p(META)	p(GeM)	Pathways
4437	MSH3	5	79950467	80172634	2.94E-08	9.52E-04	8.88E-11	1.98E-02	Repair_pathway/SSR/MMR/Mismatch_and_loop_recognition_factors
5425	POLD2	7	44154279	44163169	7.21E-04	3.12E-01	2.75E-03	5.17E-01	Repair_pathway/SSR/MMR
3978	LIG1	19	48618703	48673560	1.65E-02	8.28E-02	5.35E-04	6.39E-02	Repair_pathway/SSR/MMR
27030	MLH3	14	75480467	75518235	1.69E-02	6.69E-01	1.47E-01	6.39E-03	Repair_pathway/SSR/MMR
5395	PMS2	7	6012870	6048737	2.58E-02	3.66E-01	8.84E-03	1.76E-05	Repair_pathway/SSR/MMR
4439	MSH5	6	31707725	31730455	4.35E-02	8.54E-01	7.73E-01	5.11E-01	Repair_pathway/SSR/MMR
5982	RFC2	7	73645832	73668738	4.80E-02	5.91E-01	2.02E-02	4.44E-01	Repair_pathway/SSR/MMR
6119	RPA3	7	7676575	7758238	6.55E-02	7.22E-01	9.17E-02	4.37E-01	Repair_pathway/SSR/MMR
4292	MLH1	3	37034841	37092337	6.98E-02	3.97E-04	1.28E-04	3.91E-04	Repair_pathway/SSR/MMR

Supplementary Table 15: Summary of missing data in REGISTRY

Variable	N	Missing Values	
		Count	Percent
Motor	1744	91	4.96
Verbal Fluency	1145	690	37.6
Stroop Color	1052	783	42.67
Stroop Color	1116	719	39.18
Stroop Word	1104	731	39.84
Stroop Interference	1092	743	40.49
TFC	1758	77	4.2
FAS score	1616	219	11.93

Supplementary Table 16: Parameter estimates of variables in the model used to generate the REGISTRY cross sectional severity score. Multiple imputation adjusted estimates of statistical significance are given. CPO_1: clinical probability of onset; CPO_2: single transformation of clinical probability of onset. DF: degrees of freedom.

Parameter Estimates								
Parameter	**gender**	**Estimate**	**Std Error**	**95% Confidence Limits**		**DF**	**t for H0:**	**P Val**
Intercept		2.075589	0.267283	1.55102	2.60016	897.01	7.77	<.0001
cpo_1		-0.9142	0.21009	-1.32638	-0.50201	1191.6	-4.35	<.0001
cpo_2		-7.00283	0.911001	-8.79025	-5.2154	1141.5	-7.69	<.0001
cag		-0.01919	0.005133	-0.02927	-0.00912	862.96	-3.74	0.0002
gender	F	-0.13631	0.042605	-0.21992	-0.05271	1030.1	-3.2	0.0014
gender	M	0	0	.	.	.	.	.

Supplementary Table 17: Proportion of variance among variables present in TRACK-HD and REGISTRY which are accounted for by the first PC in the combined analysis.

Factor Pattern	
	Factor1
sqrtMotRaw	-0.91567
SDMT_correct	0.90797
SWR_correct	0.87904
tfc	0.86045

Supplementary Table 18: Effect of removing MSH3 on the Setscreen enrichment p-values for the top 14 GeM pathways in TRACK-HD, REGISTRY and the TRACK-REGISTRY meta-analysis.

Pathway	p(TRACK)	p(TRACKnoMSH3)	p(REGISTRY)	p(REGISTRYnoMSH3)	p(META)	p(METAnoMSH3)	Description
GO:32300	3.455E-09	0.04127	0.0008336	0.07162	1.13E-11	0.001024	mismatch repair complex
KEGG 3430	2.794E-07	0.04521	0.04795	0.1471	1.34E-16	0.000107	KEGG_MISMATCH_REPAIR
GO:30983	6.661E-07	0.1001	0.0004195	0.009264	3.17E-11	0.000274	mismatched DNA binding
GO:6298	0.000003533	0.2446	0.04589	0.1839	6.54E-09	0.0729	mismatch repair
GO:32407	0.01818	0.01818	0.1101	0.1101	0.000640	0.000640	MutSalpha complex binding
GO:32389	0.02249	0.02249	0.04688	0.04688	0.000523	0.000523	MutLalpha complex
GO:33683	0.08014	0.08014	0.0005874	0.0005874	0.00675	0.00675	nucleotide-excision repair, DNA incision
GO:90141	0.3318	0.3318	0.05934	0.05934	0.7872	0.7872	positive regulation of mitochondrial fission
GO:1900063	0.4103	0.4103	0.7287	0.7287	0.6926	0.6926	regulation of peroxisome organization
GO:90200	0.4582	0.4582	0.544	0.544	0.5280	0.5280	positive regulation of release of cytochrome c from mitochondria
GO:90140	0.5385	0.5385	0.3316	0.3316	0.8098	0.8098	regulation of mitochondrial fission
GO:10822	0.621	0.6228	0.6276	0.6276	0.8527	0.8527	positive regulation of mitochondrion organization
GO:4748	0.9639	0.9639	0.6974	0.6974	0.9792	0.9792	ribonucleoside-diphosphate reductase activity, thioredoxin disulfide as acceptor
GO:16728	0.9639	0.9639	0.6974	0.6974	0.9792	0.9792	oxidoreductase activity, acting on CH or CH2 groups, disulfide as acceptor

Supplementary Table 19: Effect of removing MSH3 on the Setscreen enrichment p-values for the Pearl et al. (2015) pathways in TRACK-HD, REGISTRY and the TRACK-REGISTRY meta-analysis.

Gene Set	p(TRACK)	p(TRACKnoMSH3)	p(REGISTRY)	p(REGISTRYnoMSH3)	p(META)	p(METAnoMSH3)	Description1	Description2	Description3	Description4
2071015	9.051E-07	0.3308	0.00443	0.2821	2.93E-11	0.5436	Repair_pathway	SSR	MMR	Mismatch_and_loop_recognition_factors
2071000	0.00000243	0.08225	0.06854	0.2285	1.49E-14	0.000127	Repair_pathway	SSR	MMR	
2070000	0.005767	0.2506	0.04762	0.1713	3.32E-07	0.0549	Repair_pathway	SSR		
2071017	0.01947	0.01947	0.02442	0.02442	5.84E-05	5.84E-05	Repair_pathway	SSR	MMR	MutL_homologs
2111513	0.04707	0.04707	0.2549	0.2549	0.8123	0.8123	Repair_pathway	Associated_process	TLS	DNA_polymerases
2070600	0.05024	0.05024	0.7989	0.7989	0.1098	0.1098	Repair_pathway	SSR	NER	

2070607	0.05177	0.05177	0.7606	0.7606	0.0302	0.0302	Repair_pathway	SSR	NER	TCR_(Transcription_coupled_repair)
2071104	0.05345	0.05345	0.3895	0.3895	0.0207	0.0207	Repair_pathway	SSR	BER	LONG_PATCH-BER_factors
2022100	0.0669	0.0669	0.03188	0.03188	0.00072	0.00072	Repair_pathway	DSR	Alt-NHEJ	
1100000	0.07519	0.07519	0.6138	0.6138	0.1939	0.1939	Associated_process	DNA_replication		
1080700	0.08987	0.08987	0.8346	0.8346	0.2817	0.2817	Associated_process	Checkpoint_factors	S-CC_phase	
1051930	0.1015	0.1015	0.5677	0.5677	0.1303	0.1303	Associated_process	Ubiquitin_response	Ubiquitin-_conjugating_enzymes_(E2)	UBL-conjugating_enzymes
2000000	0.1126	0.4184	0.2602	0.3906	0.0010	0.2586	Repair_pathway			
2070605	0.1144	0.1144	0.4998	0.4998	0.8140	0.8140	Repair_pathway	SSR	NER	DNA_polymerase_epsilon
1030000	0.1588	0.1588	0.1897	0.1897	0.3588	0.3588	Associated_process	Telomere_maintenance		
2070606	0.1596	0.1596	0.9556	0.9556	0.6550	0.6550	Repair_pathway	SSR	NER	DNA_polymerase_kappa
2071020	0.1726	0.1726	0.3142	0.3142	0.0099	0.0099	Repair_pathway	SSR	MMR	Other_MMR_factors
1051900	0.1973	0.1973	0.7689	0.7689	0.1711	0.1711	Associated_process	Ubiquitin_response	Ubiquitin-_conjugating_enzymes_(E2)	
2071023	0.2149	0.2149	0.1725	0.1725	0.0767	0.0767	Repair_pathway	SSR	MMR	RPA_(replication_factor_A)
1081300	0.215	0.215	0.8705	0.8705	0.4249	0.4249	Associated_process	Checkpoint_factors	HRAD17(Rad24)-_RFC_complex	
1051208	0.2409	0.2409	0.25	0.25	0.3120	0.3120	Associated_process	Ubiquitin_response	Ubiquitin_ligases_(E3)	single_Ring-finger_type_E3
1080900	0.2499	0.2499	0.4774	0.4774	0.9412	0.9412	Associated_process	Checkpoint_factors	G1-S_checkpoint	
2071003	0.258	0.258	0.8678	0.8678	0.3397	0.3397	Repair_pathway	SSR	MMR	DNA_polymerase_delta
1051222	0.2873	0.2873	0.2823	0.2823	0.1495	0.1495	Associated_process	Ubiquitin_response	Ubiquitin_ligases_(E3)	Riddle_syndrome!
1080800	0.2874	0.2874	0.3878	0.3878	0.7688	0.7688	Associated_process	Checkpoint_factors	G1-CC_phase	
2070603	0.292	0.292	0.8344	0.8344	0.5370	0.5370	Repair_pathway	SSR	NER	DNA_polymerase_delta
2071010	0.2921	0.2921	0.7597	0.7597	0.6366	0.6366	Repair_pathway	SSR	MMR	RFC_(replication_factor_C)
1051221	0.3184	0.3184	0.1559	0.1559	0.0106	0.0106	Associated_process	Ubiquitin_response	Ubiquitin_ligases_(E3)	Other_single_Ring-_finger_type_E3
1010000	0.3225	0.3225	0.4385	0.4385	0.3231	0.3231	Associated_process	Chromatin_remodelling		
1051829	0.3284	0.3284	0.5913	0.5913	0.5578	0.5578	Associated_process	Ubiquitin_response	Ubiquitin-_activating_enzymes_(E1)	UBL-activating_enzymes
1051800	0.329	0.329	0.5913	0.5913	0.5578	0.5578	Associated_process	Ubiquitin_response	Ubiquitin-_activating_enzymes_(E1)	
1051927	0.3313	0.3313	0.7885	0.7885	0.4152	0.4152	Associated_process	Ubiquitin_response	Ubiquitin-_conjugating_enzymes_(E2)	Ubiquitin-conjugating_enzymes
3060000	0.3405	0.3405	0.1703	0.1703	0.3608	0.3608	Genes_with_probable_DDR_role	Direct_Repair_(not_in_humans)		
1031600	0.3856	0.3856	0.8438	0.8438	0.5119	0.5119	Associated_process	Telomere_maintenance	Alternative_mechanism	
1031616	0.3856	0.3856	0.8438	0.8438	0.5119	0.5119	Associated_process	Telomere_maintenance	Alternative_mechanism	MRN_Complex
2020200	0.4086	0.4086	0.6981	0.6981	0.5004	0.5004	Repair_pathway	DSR	HR_(Homologous_Recombination)	

1052000	0.42	0.42	0.3114	0.3114	0.4350	0.4350	Associated_process	Ubiquitin_response	Ubiquitins_and_Ubiquitin-like_proteins	
1052028	0.42	0.42	0.3114	0.3114	0.4350	0.4350	Associated_process	Ubiquitin_response	Ubiquitins_and_Ubiquitin-like_proteins	Ubiquitins
1000000	0.426	0.426	0.4378	0.4378	0.5759	0.5759	Associated_process			
1082500	0.4288	0.4288	0.1787	0.1787	0.5650	0.5650	Associated_process	Checkpoint_factors	FPC_(fork_protection_complex)	
2111531	0.43	0.43	0.2914	0.2914	0.2994	0.2994	Repair_pathway	Associated_process	TLS	Y-family_DNA_polymerases
2071018	0.4438	0.4438	0.2644	0.2644	0.2335	0.2335	Repair_pathway	SSR	MMR	MutS_homologs_specialized_for_meiosis
2110000	0.4479	0.4479	0.4485	0.4485	0.5960	0.5960	Repair_pathway	Associated_process		
2111500	0.4479	0.4479	0.4485	0.4485	0.5960	0.5960	Repair_pathway	Associated_process	TLS	
2020000	0.471	0.471	0.4388	0.4388	0.0835	0.0835	Repair_pathway	DSR		
1050500	0.4757	0.4757	0.8548	0.8548	0.8561	0.8561	Associated_process	Ubiquitin_response	Deubiquitinating_enzyme_(DUB)	
1050501	0.4757	0.4757	0.8548	0.8548	0.8561	0.8561	Associated_process	Ubiquitin_response	Deubiquitinating_enzyme_(DUB)	UBL-specific_proteases_(ULPs)
1080000	0.4863	0.4863	0.4497	0.4497	0.8204	0.8204	Associated_process	Checkpoint_factors		
2072800	0.4971	0.4971	0.5818	0.5818	0.0702	0.0702	Repair_pathway	SSR	Other_SSR_genes	
2020400	0.5069	0.5069	0.7838	0.7838	0.8179	0.8179	Repair_pathway	DSR	NHEJ	
2071100	0.5175	0.5175	0.1144	0.1144	0.2760	0.2760	Repair_pathway	SSR	BER	
1082600	0.5196	0.5196	0.5642	0.5642	0.6168	0.6168	Associated_process	Checkpoint_factors	G2-CC_phase	
1090000	0.5699	0.5699	0.567	0.567	0.6151	0.6151	Associated_process	p53_pathway		
1050000	0.5879	0.5879	0.3435	0.3435	0.2168	0.2168	Associated_process	Ubiquitin_response		
2070602	0.593	0.593	0.1607	0.1607	0.3081	0.3081	Repair_pathway	SSR	NER	GGR_(Global_genome_repair)
2020300	0.6054	0.6054	0.5235	0.5235	0.6240	0.6240	Repair_pathway	DSR	Other_DSR_genes	
2071119	0.6093	0.6093	0.06716	0.06716	0.9067	0.9067	Repair_pathway	SSR	BER	Other_BER_factors
2071111	0.6105	0.6105	0.2266	0.2266	0.5242	0.5242	Repair_pathway	SSR	BER	AP_endonucleases
1082700	0.6144	0.6144	0.6852	0.6852	0.9253	0.9253	Associated_process	Checkpoint_factors	G2-M_checkpoint	
2021400	0.6216	0.6216	0.04964	0.04964	0.1448	0.1448	Repair_pathway	DSR	HR_(HomologousRecombination)	
1051700	0.642	0.642	0.461	0.461	0.5626	0.5626	Associated_process	Ubiquitin_response	Ubiquitin-like_proteins_(UBLs)	
1051725	0.642	0.642	0.461	0.461	0.5626	0.5626	Associated_process	Ubiquitin_response	Ubiquitin-like_proteins_(UBLs)	SUMO
1051200	0.6607	0.6607	0.07437	0.07437	0.0558	0.0558	Associated_process	Ubiquitin_response	Ubiquitin_ligases_(E3)	
1082900	0.6626	0.6626	0.8717	0.8717	0.8865	0.8865	Associated_process	Checkpoint_factors	Rad17-Mec3-_Ddc1_complex	
1082200	0.6692	0.6692	0.08041	0.08041	0.2304	0.2304	Associated_process	Checkpoint_factors	damage_in_S_phase	
2111514	0.7197	0.7197	0.5245	0.5245	0.7104	0.7104	Repair_pathway	Associated_process	TLS	epistasis_group

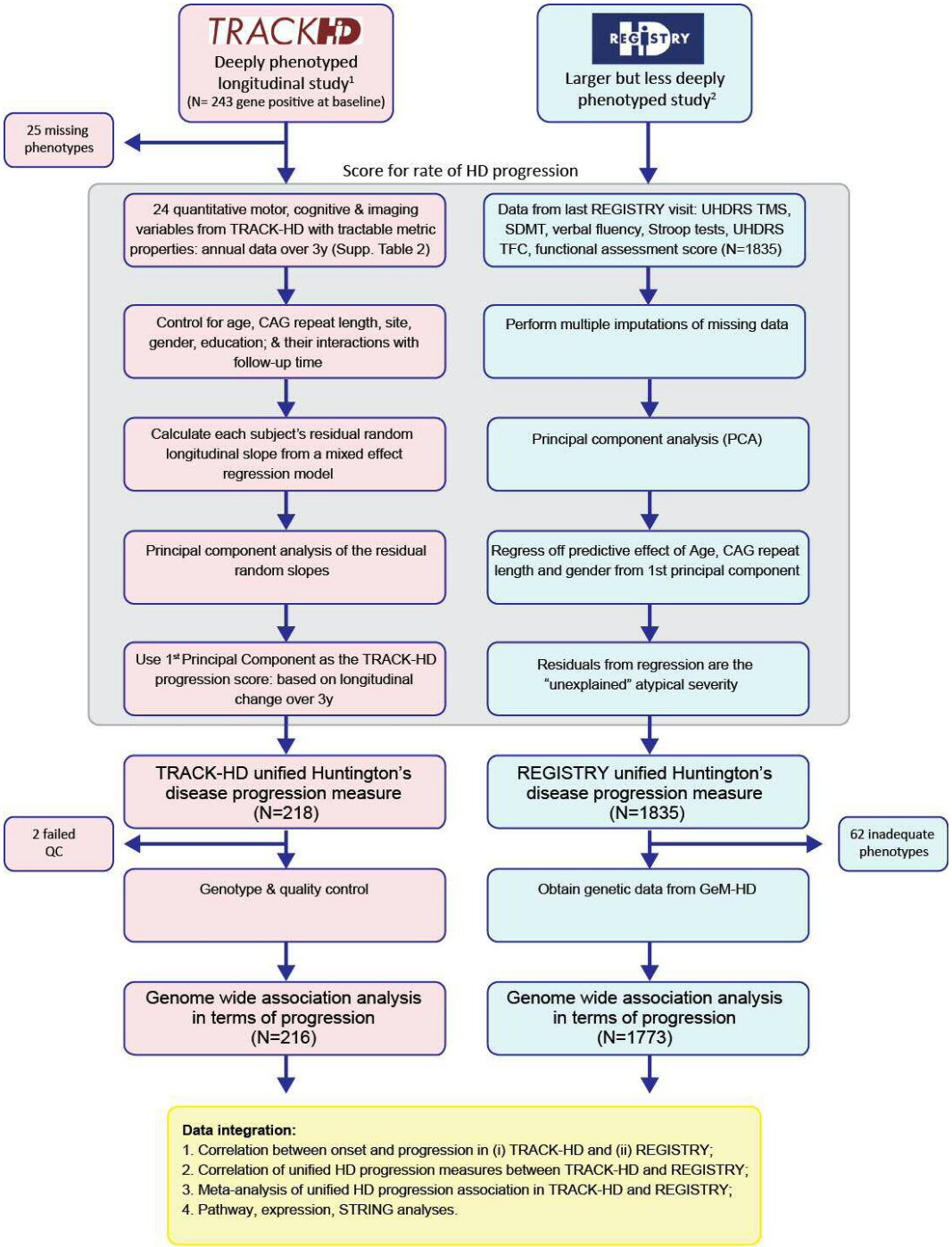
2020100	0.7228	0.7228	0.5406	0.5406	0.5703	0.5703	Repair_pathway	DSR	FA_(Fanconi_anemia_pathway)	
1040000	0.7462	0.7462	0.5933	0.5933	0.6618	0.6618	Associated_process	Chromosome_segregation		
3000000	0.7855	0.7855	0.6186	0.6186	0.3003	0.3003	Genes_with_probable_DDR_role			
2072300	0.7965	0.7965	0.3243	0.3243	0.8883	0.8883	Repair_pathway	SSR	Direct_Repair	
2072400	0.8269	0.8269	0.00389	0.003891	0.0687	0.0687	Repair_pathway	SSR	DNA_replication	
2071124	0.8385	0.8385	0.894	0.894	0.8192	0.8192	Repair_pathway	SSR	BER	SHORT_PATCH-BER_factors
2071112	0.9015	0.9015	0.1669	0.1669	0.3575	0.3575	Repair_pathway	SSR	BER	DNA_glycosylases
1051209	0.9247	0.9247	0.01381	0.01381	0.0559	0.0559	Associated_process	Ubiquitin_response	Ubiquitin_ligases_(E3)	single_Ring-finger_type_E4
1120000	0.9579	0.9579	0.6229	0.6229	0.9969	0.9969	Associated_process	Modulation_of_nucleotide_pools		
1083000	0.9619	0.9619	0.7832	0.7832	0.9161	0.9161	Associated_process	Checkpoint_factors	RAD9-Hus1-Rad1_complex	

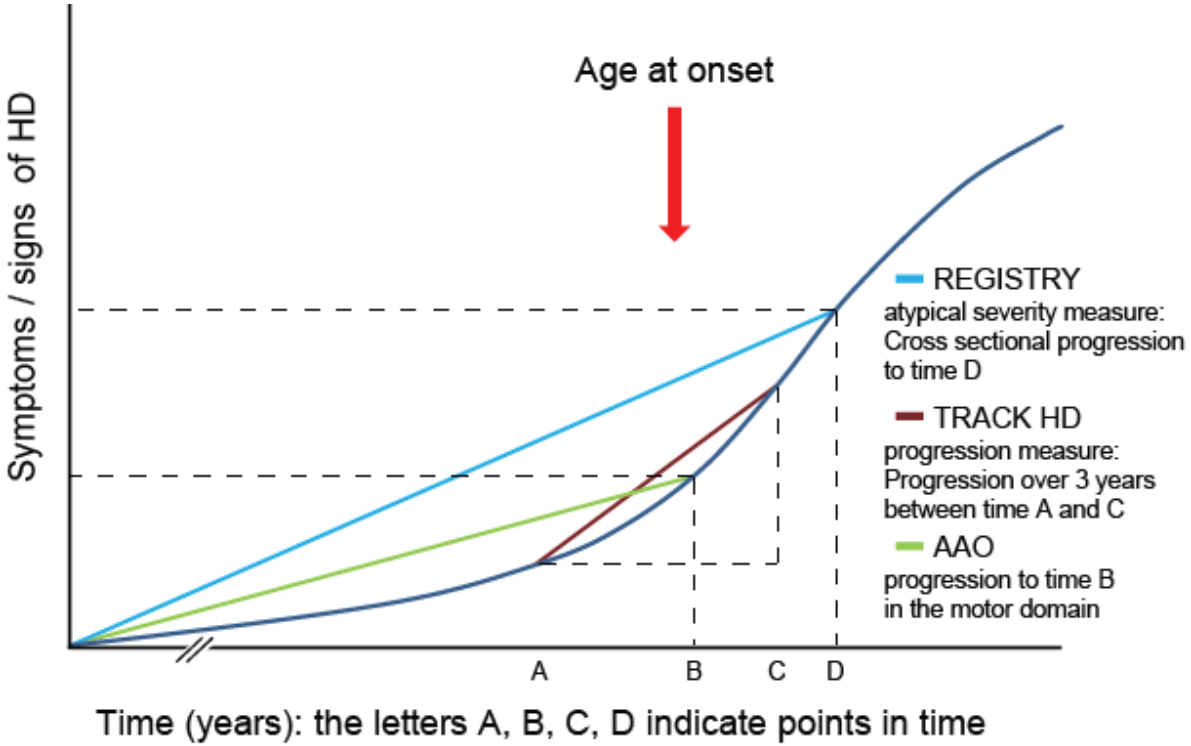
FIGURES
Identification of genetic variants associated with Huntington's disease progression: a genome-wide association study

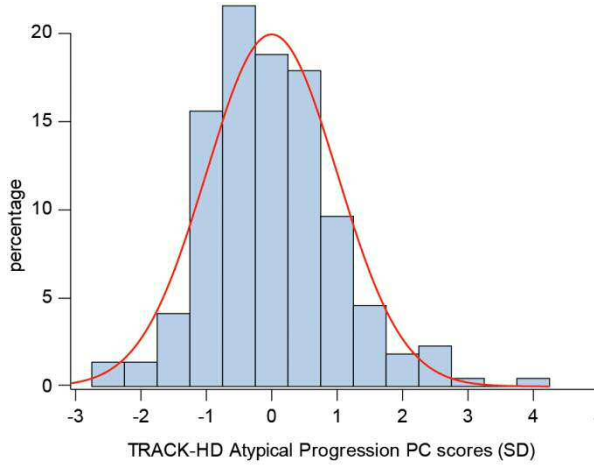
Davina J Hensman Moss*, MBBS, Antonio F. Pardiñas*, PhD, Prof Douglas Langbehn, PhD, Kitty Lo, PhD, Prof Blair R. Leavitt, MD,CM, Prof Raymund Roos, MD, Prof Alexandra Durr, MD, Prof Simon Mead, PhD, the REGISTRY investigators and the TRACK-HD investigators, Prof Peter Holmans, PhD, Prof Lesley Jones§, PhD, Prof Sarah J Tabrizi§, PhD

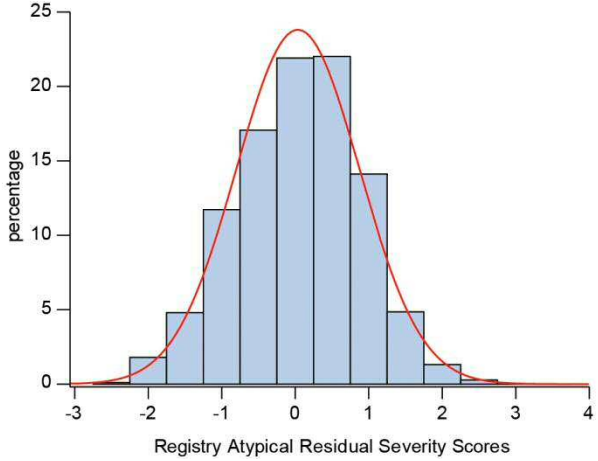
* These authors contributed equally to this work

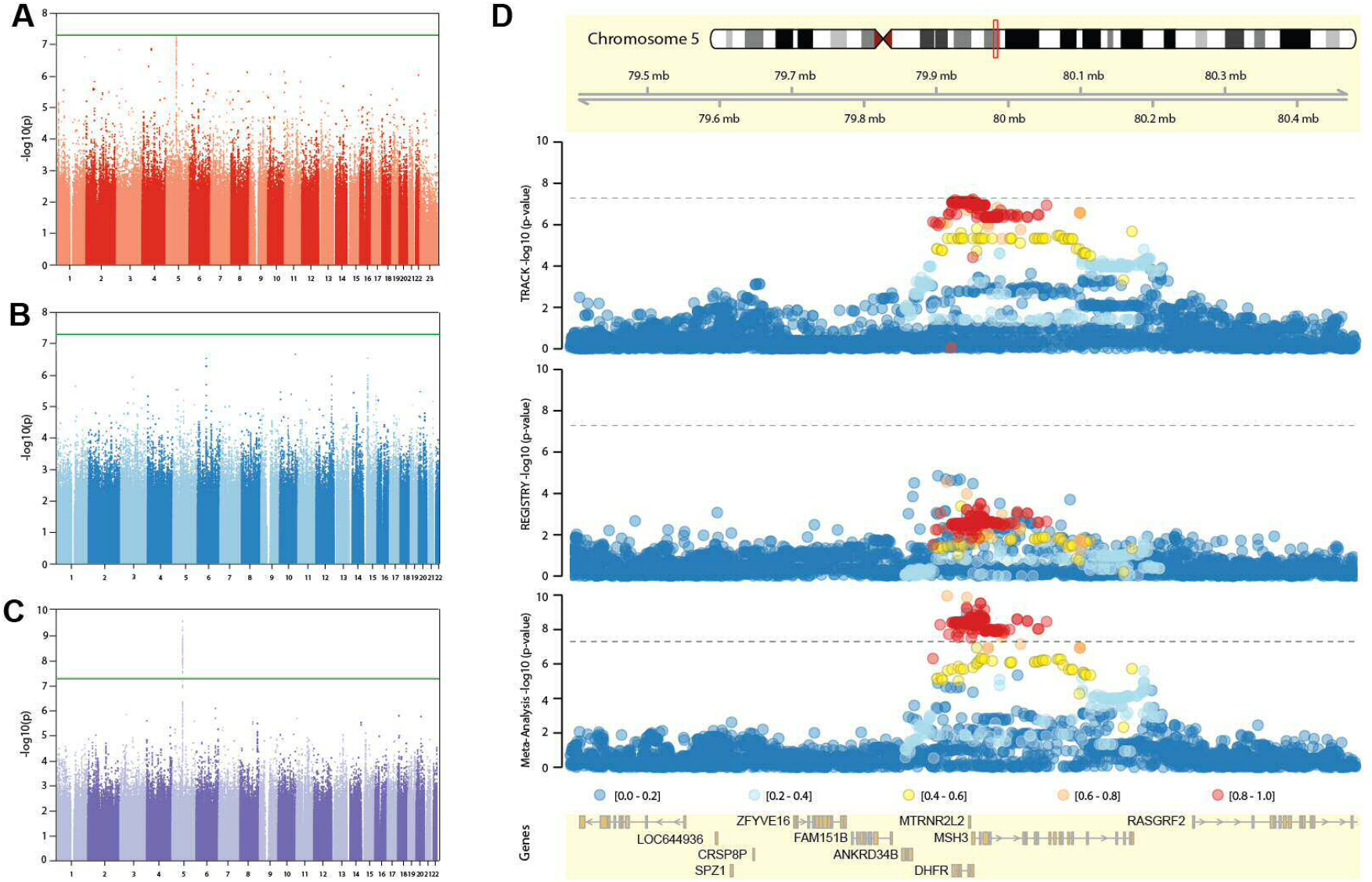
§ These authors contributed equally to this work

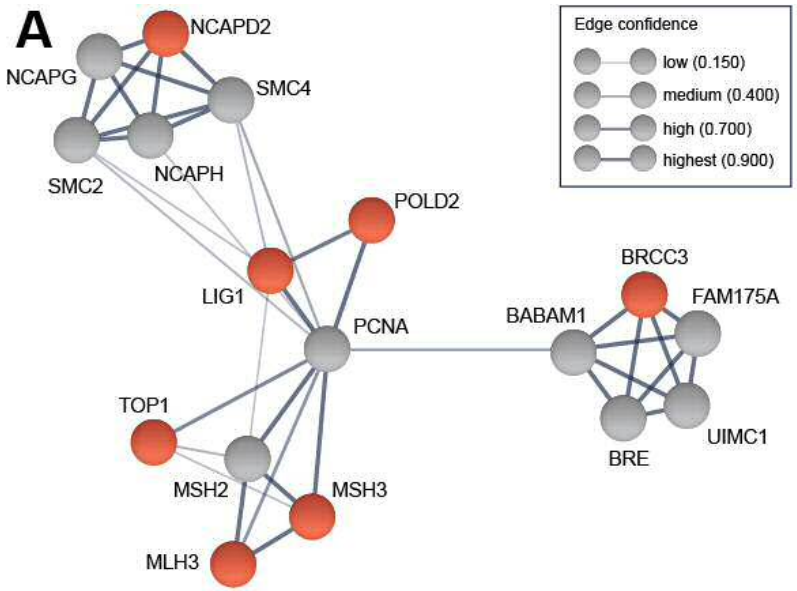

TRACKHD
Deeply phenotyped longitudinal study[1]
(N= 243 gene positive at baseline)
REGISTRY HD
Larger but less deeply phenotyped study[2]
25 missing phenotypes
Score for rate of HD progression
24 quantitative motor, cognitive & imaging variables from TRACK-HD with tractable metric properties: annual data over 3y (Supp. Table 2)
Data from last REGISTRY visit: UHDRS TMS, SDMT, verbal fluency, Stroop tests, UHDRS TFC, functional assessment score (N=1835)
Control for age, CAG repeat length, site, gender, education; & their interactions with follow-up time
Perform multiple imputations of missing data
Calculate each subject's residual random longitudinal slope from a mixed effect regression model
Principal component analysis (PCA)
Principal component analysis of the residual random slopes
Regress off predictive effect of Age, CAG repeat length and gender from 1st principal component
Use 1st Principal Component as the TRACK-HD progression score: based on longitudinal change over 3y
Residuals from regression are the "unexplained" atypical severity
TRACK-HD unified Huntington's disease progression measure (N=218)
REGISTRY unified Huntington's disease progression measure (N=1835)
2 failed QC
62 inadequate phenotypes
Genotype & quality control
Obtain genetic data from GeM-HD
Genome wide association analysis in terms of progression (N=216)
Genome wide association analysis in terms of progression (N=1773)
Data integration:
1. Correlation between onset and progression in (i) TRACK-HD and (ii) REGISTRY;
2. Correlation of unified HD progression measures between TRACK-HD and REGISTRY;
3. Meta-analysis of unified HD progression association in TRACK-HD and REGISTRY;
4. Pathway, expression, STRING analyses.

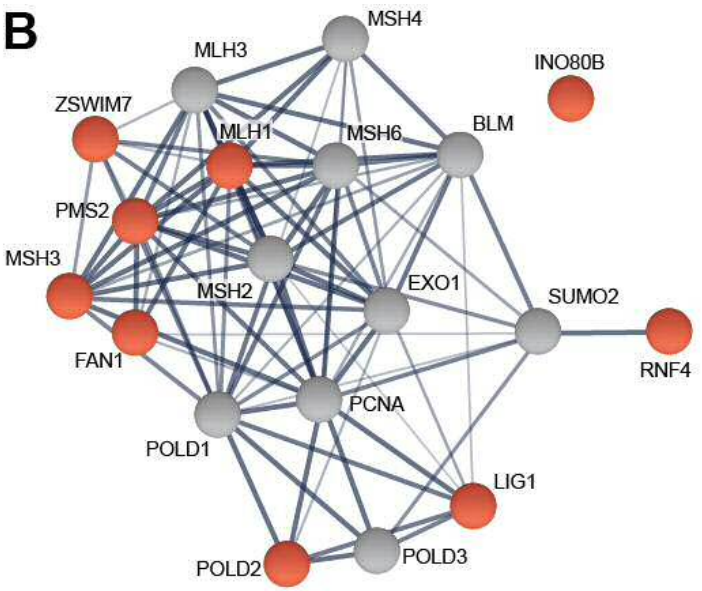

Age at onset
Symptoms / signs of HD
REGISTRY
atypical severity measure:
Cross sectional progression
to time D
TRACK HD
progression measure:
Progression over 3 years
between time A and C
AAO
progression to time B
in the motor domain
A
B
C
D
Time (years): the letters A, B, C, D indicate points in time

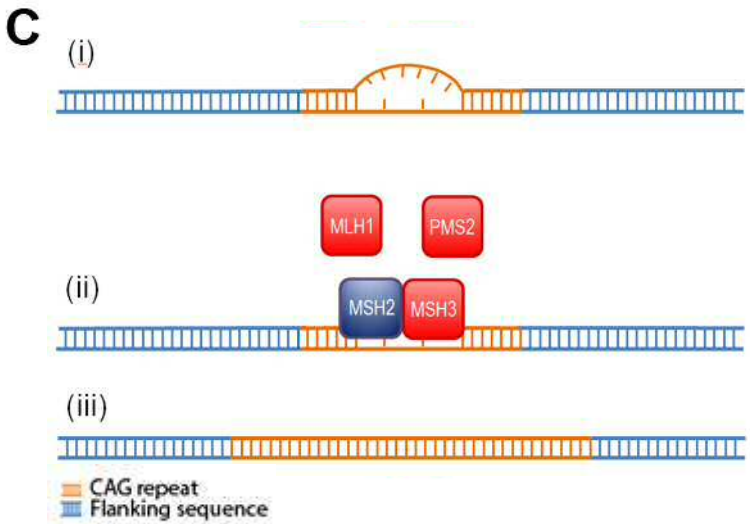

percentage
20
15
10
5
0
-3
-2
-1
0
1
2
3
4
TRACK-HD Atypical Progression PC scores (SD)

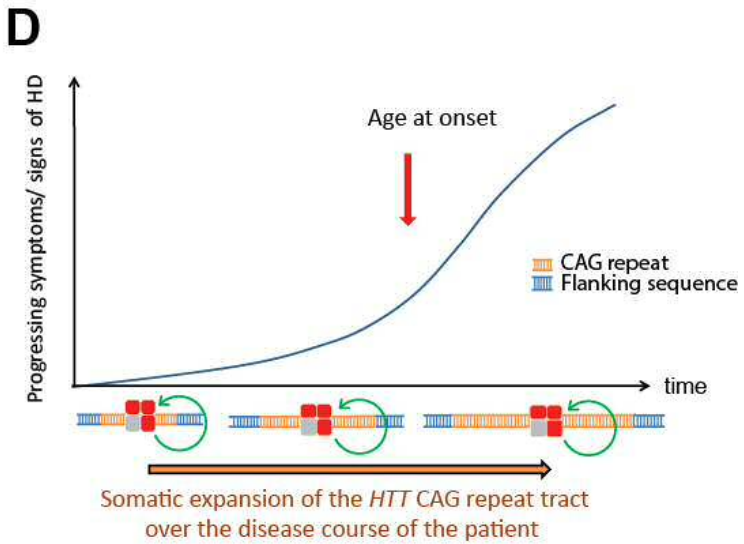

percentage
25
20
15
10
5
0
-3
-2
-1
0
1
2
3
4
Registry Atypical Residual Severity Scores


A
B
C
D
-log10(p)
Chromosome 5
79.5 mb
79.7 mb
79.9 mb
80.1 mb
80.3 mb
79.6 mb
79.8 mb
80 mb
80.2 mb
80.4 mb
TRACK -log10 (p-value)
REGISTRY -log10 (p-value)
Meta-Analysis -log10 (p-value)
[0.0 - 0.2]
[0.2 - 0.4]
[0.4 - 0.6]
[0.6 - 0.8]
[0.8 - 1.0]
Genes
ZFYVE16
MTRNR2L2
RASGRF2
LOC644936
FAM151B
MSH3
CRSP8P
ANKRD34B
SPZ1
DHFR


A
NCAPD2
NCAPG
SMC4
SMC2
NCAPH
POLD2
LIG1
PCNA
BRCC3
FAM175A
BABAM1
UIMC1
BRE
TOP1
MSH2
MSH3
MLH3
Edge confidence
low (0.150)
medium (0.400)
high (0.700)
highest (0.900)


B
MSH4
MLH3
INO80B
ZSWIM7
MLH1
MSH6
BLM
PMS2
MSH3
MSH2
EXO1
SUMO2
FAN1
RNF4
PCNA
POLD1
LIG1
POLD2
POLD3


C
(i)
MLH1
PMS2
(ii)
MSH2
MSH3
(iii)
CAG repeat
Flanking sequence


D
Progressing symptoms/ signs of HD
Age at onset
CAG repeat
Flanking sequence
time
Somatic expansion of the HTT CAG repeat tract
over the disease course of the patient